AD NUMBER
AD000191
NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 07 NOV 1952. Other requests shall be referred to Army Adjutant Generals Office, Washington, DC 20310.

AUTHORITY

usapro ltr 15 Aug 1966

rmed Services Technical Information Ag

DOCUMENT SERVICE CENTE

KNOTT BUILDING, DAYTON, 2, OHIO

AD —

191

UNCLASSIFIED

PRS REPORT 976

DEVELOPMENT OF ARMED FORCES

QUALIFICATION TEST AND PREDECESSOR

ARMY SCREENING TESTS, 1946-1950

PERSONNEL

RESEARCH

SECTION

DEPARTMENT
OF THE ARMY

PR & PBR-PERS BUR-TAGO

Personnel Research Section
Personnel Research and Procedures Branch
The Adjutant General's Office
DEPARTMENT OF THE ARMY

SENIOR STAFF

Dr. Donald E. Bales, Chief, PRS

Dr. Edward A. Rundquist, Research Manager

Dr. Hubert E. Brogden, Research Adviser

Dr. Richard H. Gaylord, Chief, Research Operations I

Dr. Julius E. Uhlaner, Chief, Research Operations II

Mr. Harry H. Harman, Chief, Research Services

The Office of the Personnel Research Section is located at Wake Building, Oklahoma Avenue and E Streets, N. E. Washington, D.C.

036970

PRS REPORT 976

DEVELOPMENT OF ARMED FORCES QUALIFICATION TEST AND PREDECESSOR ARMY SCREENING TESTS, 1946-1950

J. E. Uhlaner, and Daniel J. Bolanovich

Approved by J. E. Uhlaner,
Program Coordinator

7 November 1952

FOREWORD

DEVELOPMENT OF THE ARMED FORCES QUALIFICATION TEST AND PREDECESSOR ARMY SCREENING TESTS, 1946-1950
(Based on PRS Report 976)

The Armed Forces Qualification Test (AFQT) is the means of determining mental test acceptability of potential enlistees and inductees. It is jointly developed by the Armed Forces to implement the mental standards established by law and administrative action for admission into the Armed Forces. The AFQT also provides a basis for qualitative distribution of manpower among the Services.

This report is a summary of the work done and the problems encountered in the development and use of the AFQT. It describes briefly the experience with other tests used for initial screening and related purposes in the past. It was out of this experience that the current AFQT was developed.

The test was developed in two comparable forms in accordance with accepted principles and techniques of test construction and with due regard for policy and operating problems. When administered according to prescribed procedures, the AFQT adequately maintains its effectiveness as a mental test screen. Continual research is underway to maintain and improve mental test screening procedures.

This report is of interest to research technicians, and to those responsible for or concerned with initial screening in the Armed Forces.

DEVELOPMENT OF ARMED FORCES QUALIFICATION TEST AND PREDECESSOR ARMY SCREENING TESTS, 1946-1950

CONTENTS

TABLES

FIGURES

DEVELOPMENT OF THE ARMED FORCES QUALIFICATION TEST AND PREDECESSOR ARMY SCREENING TESTS, 1946-1950

INTRODUCTION

Mental test standards are established by law and military policy as a means of determining acceptability of potential enlistees and inductees. The purpose of such standards is to screen out those who cannot profit from military training and who might be actual liabilities to the Services. Such standards must vary from time to time in accordance with changes in manpower demands, training facilities, and National Policy.

Psychological screening tests are used to examine potential enlistees and inductees for acceptance in conformity with mental standards. Similar tests were used during World War II for classification purposes. Experience with classification tests indicated their value for initial screening. Constant improvement in the tests and their use is necessary because of obsolescence and the necessity to refine sensitivity at various cut-off points as these are changed by law or policy.

The program described in this report led to the development of the tests currently used (October 1952) for initial screening. These tests, Armed Forces Qualification Test AFQT-1, and Armed Forces Qualification Test AFQT-2, were based on extensive research with earlier forms of screening tests and on follow-up studies to determine the applicability of AFQT to operating conditions. The studies relating to the development and use of AFQT represented the joint effort by Army, Navy, Marine Corps, and Air Force research personnel by direction of the Department of Defense. The Army was designated as the coordinating and executive agency in this program. The earlier tests had been developed for Army use only.

The purpose of this report is to summarize the problems encountered in the development of AFQT and to describe the outcome of attempts to solve these problems. Many of these problems arose in the development and use of earlier tests. To permit a better understanding of the development of AFQT, a summary of experience with earlier tests is provided as a background.

BACKGROUND

CLASSIFICATION VS. SCREENING TESTS

Because much of the history of the screening tests developed under this program is related to previous and concurrent developments of classification tests, and because similar score conversion systems and terminology frequently apply to both, it is important to clarify the difference between these two types of tests.

Classification Tests are used primarily to classify men on the basis of abilities for assignment such as officer training and specialist training. These tests originally measured just a few abilities (in most cases similar to those currently measured by screening tests), and in the Army were referred to as "Army General Classification Tests," or AGCT tests. More recently the classification tests have been expanded to a battery of specific ability measures called the "Army Classification Battery." Classification tests are administered to men at reception centers after they have been accepted for service in the Army.

Screening Tests are tests used at recruiting or induction stations for determination of mental fitness for service in the Army. Cut-off scores are used on screening tests to determine whether an applicant or selective service registrant will be accepted or rejected for service insofar as mental qualifications are concerned. Tests labeled AGCT are not used in screening, although items from the AGCT tests have been used to comprise screening tests, and in one case two AGCT tests were given different titles and used for screening purposes.

ARMY GENERAL CLASSIFICATION TESTS (AGCT SERIES)

The Army General Classification Tests (AGCT series) were introduced in 1940. During the period 1940-1949 they were gradually replaced by the Army Classification Battery. The development of both of these tests set precedents with respect to standardization methods and item types which affected the construction, standardization, and application of screening tests developed under this program. For that reason, it is well to review briefly pertinent aspects of the history of AGCT tests. These are reviewed below in chronological order of their appearance.

AGCT-1a: This test consisted of vocabulary, arithmetic reasoning, and block counting items in equal numbers to make a total of 150 test items. Items were of the multiple-choice type with four alternatives, and the test was scored rights minus one-third wrongs to yield a single raw score. This scoring formula was used because at the time it was believed that a correction for chance was justified. The test was standardized on a population of CCC enrollees and soldiers, all white, and between the ages of 20 through 29 (N = 2675). These men were divided into stratified cells based on combinations of age, highest school grade reached, and geographical area. Then, in order to establish norms representative of the total US male population between the ages 20 through 29, AGCT scores for these men in the various stratified cells were weighted to provide representation proportional to that of such age-education-geographical groups in the 1930 census. The distribution of scores so obtained was adjusted for convenience to a standard score scale of Mean = 100 and Standard Deviation = 20. Such equivalent standard scores were obtained for each AGCT-1a raw score. These converted scores became known as "Army Standard Scores." The Army Standard Score system has been applied to practically all subsequent classification and screening instruments.

To meet operational requirements, mental test scores were grouped broadly on the basis of Army Standard Scores into five "Army Grades" or mental groups. Originally they were as shown in Table 1.

Table 1. Mental group classifications of Army Standard Scores.

Army Grade (Mental Group)	Army Standard Score Range*
I	130 and higher
II	110 - 129
III	90 - 109
IV	70 - 89
V	69 and lower

* Percentages of the Army Population falling in each mental group
  varied from time to time with changes in norms. These percentages
  for current norms are shown in the later section on Operational
  Application of AFQT-1 and AFQT-2.

These groups remained as defined above, except that in July 1942 the above limit of group IV was changed from standard score 70 to standard score 80. This was done so that the distribution of scores would correspond better with the distributions anticipated from operational use. Although the standard score ranges for groups IV and V varied, this grading system has remained with the Army, and equivalence of subsequent selection and classification test scores to the various grade level limits is an important aspect of their standardization. As will be seen later, this grade system in 1950 became the basis for controlling the distribution of armed forces input among the Army, Air Force, Navy, and Marine Corps. The test AGCT-1a was introduced operationally in October 1940.

AGCT-1b: This test had the same general composition as form 1a. It contained 50 arithmetic, 50 vocabulary, and 50 block counting items. The 50 block counting items were, in fact, identical to those in form 1a. It was standardized on a population of 3,856 soldiers drawn from 8 of the 9 Army Corps areas. The method used was line-of-regression equation of 1b scores to standard scores on 1a. The test was introduced in April 1941.

AGCT-1c and -1d: Each of these tests consisted of 140 items--47 vocabulary, 49 arithmetic, and 44 block counting--arranged in spiral omnibus form. These items also had four alternatives, and the tests were scored rights minus one-third wrong, as in AGCT-1a and -1b.

Forms 1c and 1d were standardized on a population of 1,782 soldiers from two Army Corps. Scores on forms 1c and 1d were equated to Standard Scores on form 1a by a combination of the equipercentile method and the line-of-regression method. They were introduced operationally in October 1941.

AGCT-3a: This test departed from previous AGCT forms in that it was actually a battery of four tests which could be scored in total, or separately in order to provide separate measures of the abilities measured by AGCT. This was the beginning of the "Classification Battery" idea. The four component tests of AGCT-3 were (1) Reading and Vocabulary, (2) Arithmetic Reasoning, (3) Arithmetic Computation, and (4) Pattern Analysis. These tests were each of the multiple choice, four alternative answer type, and were scored rights minus one-third wrongs. The total score was the sum of raw scores on the four components.

AGCT-3a was standardized on a population of 39,178 soldiers stratified by Service Command, color, age, and education. The method of standardization used was equipercentile equating of AGCT-3a raw scores to Standard Scores on AGCT-1a and -1b.

This test was introduced in April 1945. As of this date, it is still used by the Marine Corps for classification.

AGCT-3b: This was an alternate form of AGCT-3a, having the same composition and item types but using different items. AGCT-3b was standardized on a group of 1,000 soldiers at Camp Atterbury selected to match proportionally the numbers of white and colored men and the distribution of AGCT-3a scores to those of the Army as a whole. Standard Scores for raw scores on AGCT-3b were obtained by equipercentile equating to Standard Scores on AGCT-3a.

This test was introduced for alternate use with AGCT-3a shortly after VJ Day, August 1945.

ARMY CLASSIFICATION BATTERY (ACB)

The AGCT tests were predominantly of the verbal type, including items of the arithmetic reasoning, vocabulary, reading comprehension, and spatial relations types. As early

as 1941, research and operating experienc· indicated the need for supplementary tests for
use in classification. Beginning in 1941, specific tests such as Mechanical Aptitude, Cleri-
cal Speed, Radio Code Learning, and Automotive Information were introduced at various
times to supplement AGCT in classification. By the fall of 1947 ten such tests had been in
use for classification purposes; but interpretations of the meaning and appropriate use of the
test scores varied widely because classification officers differed in the amount of their tech-
nical knowledge, and it was not possible at the time to make available sufficient data on validity
and interrelationships so that even technically-trained personnel could make optimum use of
the tests. These ten tests made up the <u>Army Classification Battery</u>. Work was begun on a
continuing program to study various combinations of these tests which were valid for groups
of Army MOS's. These combinations, predictive of performance for similar MOS groups,
were called "Aptitude Areas." By the spring of 1949, the ten classification tests were grouped
into ten "Aptitude Areas." At this time, the Army Classification Battery, making use of the
"Aptitude Area" system for classification at Reception Centers, was introduced officially for
classification of soldiers. The tests AGCT-3a and -3b were withdrawn. It should be noted,
however, that three of the subtests of 3a and 3b--Reading and Vocabulary, Arithmetic Rea-
soning, and Pattern Analysis--were retained as separate tests in the Army Classification
Battery, and made up three of the ten classification tests in the Battery.

DEVELOPMENT AND USE OF SCREENING TESTS

The use of psychological tests during World War II for purposes of initial screening for
Service began soon after induction became effective. Wartime screening for mental abilities
passed through four phases, each characterized by psychological testing procedures.

The first phase included the wartime period prior to August 1942. Regulations at this
time excluded from military service all men who did not have the capacity for "reading and
writing the English language as commonly prescribed for the fourth grade in grammar
school." It was further prescribed that men who had not completed the fourth grade would be
tested at induction stations to determine whether they possessed this capacity. Toward the
latter part of this period a test called the "Minimum Literacy Test" was developed for this
purpose. There were 12 forms of this test with 12 simple questions each. Each form had a
passing score of 9 correct answers, which was considered equivalent to fourth grade reading
and writing ability, and set to differentiate mental group IV from mental group V. Although
6 of these 12 test forms were placed in the hands of the National Selective Service Head-
quarters for use by local boards in preliminary screening, local boards did not generally use
them. Responsibility for screening with the Minimum Literacy Test was placed with the
induction station.

The second phase included the period August 1942 to June 1943. During this period
induction of men who could not meet the above literacy standards was permitted provided they
possessed sufficient intelligence to absorb military training rapidly. The induction of such
men was limited to fixed quotas in terms of percentages of the total number of men inducted
at each station each day. These quotas varied from 10% at the outset to 5% in February 1943.
This regulation introduced to the Services men designated as <u>illiterate</u>, and was the beginning
of screening on the basis of <u>mental ability</u> in addition to literacy. These new regulations
required a more comprehensive system of screening at induction stations. On the basis of
research, a series of multiple hurdles was applied to men who did not qualify by virtue of
fourth grade education. Those who passed any one hurdle were considered mentally accepta-
ble. These hurdles consisted of: (1) The Army Information Sheet, a revision of the Minimum
Literacy Test which was used to determine acceptable literacy; (2) The Visual Classification
Test, a non-language group test of mental ability, administered to those who did not pass on
literacy; and (3) A battery of two individual mental tests, the Concrete Directions Test and the
Block Counting Test, given those who did not pass the group mental test. Those who did
not qualify in any of these tests were rejected.

- 4 -

During the third phase, Jun. 1943 to May 1944, no limits were placed on the number of "illiterates" to be inducted because of the need for rapid expansion. This phase also differed from the previous two in that special training units were set up to give these men literacy training before they were forwarded to regular training centers. For screening at this time, a new test called the Qualification Test replaced the Army Information Sheet as the initial literacy screening instrument. It consisted of series of items--number series, space orientation, arithmetic, and reading and vocabulary--arranged in cycles of increasing difficulty. The Visual Classification Test and the individually administered Concrete Directions and Block Counting tests remained in the battery as before. A passing score on any one test hurdle qualified a man mentally.

In the fourth phase, June 1944 to the end of World War II, a new series of mental tests which had been subjected to considerable test construction and validation study was introduced. Validation studies of these tests were made with full consideration of the need for distinguishing between literacy, on the one hand, and over-all performance as a soldier, on the other. This is a continuing problem in the validation of mental tests since evidence is available which indicates that these two aspects are not highly correlated with each other.

The new series of tests was incorporated in a mental screening procedure which included the following instruments: (1) The Qualification Test, which was used as in the previous phase, except that the passing score was raised in accordance with research findings and an alternate form was developed; (2) The Group Target Test, which replaced the Visual Classification Test; (3) The Individual Test; and (4) The Non-Language Individual Examination. At this time all men who failed the Qualification Test, but were accepted on passing a subsequent hurdle, were placed in Special Training Units before being assigned to Basic Training. Those who could not learn (as determined by the Specialized Training Unit) the required academic and military subjects during a maximum of 13 weeks in Special Training Units were discharged and returned to civilian life.

Screening tests during World War II were used to select men on the basis of very low mental standards, i.e., those who did not possess sufficient literacy or mental ability to absorb the most elementary training. However, this experience emphasized the value of screening on the basis of mental ability--a practice which continued after the war when higher mental standards could be applied.

CLASSIFICATION TESTING CONCEPTS OF SIGNIFICANCE TO SCREENING TESTING

Several concepts utilized in the development of classification testing in the Army had a significant bearing on research in the development of screening tests under this program. Those concepts which are most important to this program are:

The "Army Standard Score" System. The system of conversion of raw scores on tests to Army Standard Scores, which began with AGCT-1a, provides an established frame of reference for interpreting test scores. This concept has remained with the Army to the present. The "Army Standard Score" distribution was originally defined as having a Mean of 100 and a Standard Deviation of 20. Though the Army Standard Score is still the basic means of converting raw scores, it no longer has its original definition in terms of mean and standard deviation and the percentages expected from probability tables no longer apply. Through successive use of tie-back standardization procedures (to be described below) on subsequent tests, the Army Standard Score on AGCT and similar tests has come to mean the score on those tests which is equivalent to that standard score on the original AGCT-1a. Thus, for example, a score of 65 on AFQT means that an equivalent score of 65 would be obtained on AGCT-1a, regardless of the percentage of men making that score at any given time.

- 5 -

Interpretation of test scores by ref rence to Army Standard Scores has become very familiar to Army personnel and Army Standard Score conversions for tests of the AGCT type are almost a necessity. In addition, certain standard scores which have been used at various times as reference points and cutting scores have become important bench marks by which operating Army personnel evaluate mental ability.

Mental Groups. The five mental groups, I, II, III, IV, V (see Table 1), used in classifying men by mental ability have also become an established concept in the Army. These mental groups were used as a basis for allocating men within the Army organization to various units. Mental groups have been defined in equivalent terms for all AGCT tests. The concept of mental groupings was taken over later for screening tests, and some of the earlier screening tests were redesigned specifically to discriminate between mental groups III and IV (instead of between groups IV and V), as was presumed to be required for the higher standards of a peacetime Army. As will be discussed later, this system of mental groups also was used as a basis for allocating personnel among the Army, Navy, and Air Force, beginning in 1950.

Tie-Back Standardization Methods. In a previous discussion, it was pointed out that AGCT tests developed subsequent to AGCT-1a were standardized by computing raw scores on these tests equivalent to the standard scores originally developed for AGCT-1a on a group selected to represent the total US male population between the ages of 20-29 years in 1940. The standard scores then represented norms for the AGCT tests. This tie-back type standardization was accomplished by: (1) Choosing a sample population within the Army; (2) Administering the test to be standardized and a previous form of AGCT test (reference test) upon which standard scores had been established; and (3) Computing standard score equivalents in the reference test for raw scores on the test being standardized by means of equipercentile equivalents or line-of-regression equivalents. One advantage of this method is that each standardization does not require a precisely representative sample of the total population to which the test norms apply. It does require, however, that each new standardization sample contain a representation of cases at all score levels throughout the total range, and that no extraneous biasing variables be introduced in selecting the sample. Standardization by tie-back methods was used entirely in the development of screening tests under this program.

PROBLEM

ORIGIN OF THE NEED FOR SCREENING TESTS

At the close of World War II in 1945, involuntary inductions were stopped and voluntary recruitment began again for the Armed Forces. The screening tests used for induction of marginally mental and illiterate groups were dropped from use.

One test which was used specifically as a screening instrument for limited service personnel during World War II, showed promise as a screening device for recruits in peacetime. This test was called R-1(1), introduced in October 1942 to screen inductees who were limited physically but who could be used in restricted assignments. It was a short test made of 50 items from AGCT-1a (17 vocabulary, 18 arithmetic, and 15 block counting items). Items for the test had been selected so as to give maximum discrimination between men in mental grade III and those in mental grade IV. The test was standardized and calibrated against AGCT-1a to yield raw score equivalents to AGCT-1a standard scores. A relatively high cutting score of standard score 90 on R-1 was established as a requirement for induction of such personnel to compensate for the limited physicals.

In April, 1946 the test R-1 was introduced officially as a means of determining mental qualification of applicants for enlistment in the Army, and a raw score of 15 (standard score

70) was established as the minimum acceptable. The test was administered at Central Examining Stations.

At this time the procedure for processing recruits which related to mental testing was as follows: (1) Acceptance of application and initial screening on obvious basic qualifications, such as age, major physical handicaps, etc.; (2) Forwarding to Central Examining Station for final acceptance or rejection, including administration of the test R-1; and (3) Enlistment and forwarding to the appropriate base for processing.

It was realized early that much waste in costs of transportation and processing of enlistment applicants could be avoided if those who were likely to be disqualified for mental reasons could be detected and rejected at local recruiting stations before being sent to Central Examining Stations. Therefore, an important need existed for new screening tests which would serve two functions: (1) Use at Central Examining Stations for final determination of acceptance or rejection; and (2) Use at local recruiting stations by relatively untrained recruiting personnel for pre-screening applicants before referral to Central Examining Stations.

OPERATIONAL REQUIREMENTS OF SCREENING TESTS

The specific problems affecting the development of screening tests under this program varied with changes in operational policy and procedure. As occasioned by these changes and as tests became obsolete, new forms were developed. Some of the major operational requirements which influenced the nature of screening tests developed throughout this program were:

1. Level of Mental Standards prescribed for admission at various times.

2. Level of training provided for enlisted men.

3. Provisions for administration of tests, in terms of time and training of administrators.

4. Use of tests at local recruiting stations or Central Examining Stations.

5. Reintroduction of involuntary induction.

6. Adoption of uniform screening standards and instruments by all Armed Forces.

7. Use of screening instruments as a basis for allocation of personnel to the four Armed Services.

GENERAL STATEMENT OF PROBLEM

The general problem of this program was to develop screening tests for selection of personnel procured by recruitment and induction to fit existing operational requirements and to provide for continuous improvement of these instruments and related procedures.

SPECIFIC OBJECTIVES OF THE PROGRAM

Specific objectives of the program were defined in terms of the three phases which composed it, as follows:

1. Development of Recruitment Tests: R-2, R-3, R-4, R-5, and R-6. The objective of this phase was to develop screening tests which could be administered for pre-screening at local Army recruiting stations or for final mental screening at Central Examining Stations.

2. Development of Armed Forces Qualification Test (AFQT). The objective of this phase was to develop, through joint Army, Navy, and Air Force effort, a screening instrument which could be used by all three Services at their main examining stations for purposes of determining acceptance or rejection of enlistees or inductees.

3. Follow-Up of the Standardization of AFQT. The objective of this phase was to recheck the standardization of AFQT and to determine the effect of operational administration of the test upon norms and standards established on the basis of its original standardization.

Further detailed objectives of these three phases will be discussed in connection with the following report of research accomplished.

DEVELOPMENT OF RECRUITMENT TESTS: R-2, R-3, R-4, R-5, AND R-6

Each of the recruiting tests, R-2, R-3, R-4, R-5, and R-6 were first developed for use as screening devices at main examining stations. All except R-5 and R-6 were later transferred to use as prescreening instruments at local recruiting stations as forms were developed for use at main stations. Appendix A lists the chronology of introduction and use of these tests as specified in Army regulations.

DEVELOPMENT OF R-2

R-2 was originally intended to be an alternate form to R-1 for screening limited service inductees. When induction became the sole source of procurement, work on R-2 was discontinued. With the return of the Army to a peacetime basis in 1946, induction was discontinued and voluntary enlistment became the sole means of entry into the Army. New screening tests were needed, and among these, the development of R-2 was resumed for use as an alternate to R-1 administered at local recruiting stations.

In constructing R-2, 35 items were selected[2] from AGCT-1b (17 vocabulary and 18 arithmetic) which discriminated between men of mental grades III and IV. The fifteen block counting items used in R-1 were added to make an omnibus 50-item test, essentially a short form of AGCT-1b. The fifteen block counting items were retained from R-1 since AGCT-1a, of which R-1 was a short form, contained the same block counting items as AGCT-1b. The vocabulary and arithmetic items for R-2 were selected on the basis of the significance of differences in their p-values between groups of 500 mental grade III and 500 mental grade IV men from various reception centers, and on the basis of matched difficulty distribution with similar content items of the test R-1.

R-2 was standardized first in 1942[3] in order to determine "critical" score equivalents to standard scores of 90 and 100 on the AGCT-1c scale. The population used was a group of 375 men who came to the reception center at Camp Lee, Virginia, on 25 and 26 March 1942. The men represented all five mental grades on AGCT-1c. Both AGCT-1c and R-2 were given to the group using a counterbalanced order of administration. Equivalent scores on R-2 for standard scores of 90 and 100 on AGCT-1c were determined by the equipercentile method. The reliability of R-2 was estimated at .94 and its correlation with AGCT-1c at .83. The correlation between R-2 and R-1 was .87.

Later in 1946, in connection with the standardization of R-3 and R-4, the test R-2 was standardized on a group of 700 enlistees at Camp Atterbury Reception Center[4] chosen to represent proportionally the distribution of the general Army population among the five mental grades on AGCT-3a. Equivalents to standard scores on AGCT-3a for each raw score on R-2 were obtained by the equipercentile method.

DEVELOPMENT OF R-3 AND R-4

Early in 1946, work was begun on a test for use in Army enlistment which could be administered individually or in groups with a minimum of verbal instruction, could be completed in less than 30 minutes, and would provide scores as nearly equivalent to AGCT-3a standard scores as possible within the lower test score ranges.

While R-2 had not been introduced operationally at this time, it seemed probable that items selected from the residue available after constructing AGCT-3a and AGCT-3b would correlate more highly with AGCT-3a than would those selected from AGCT-1b for R-2. Furthermore, R-2 was still in need of further research. It was decided to study R-2 further, along with the development of R-3 and R-4, and to compare R-2 with the newly developed tests.

Preparation of Experimental Tests R-3x and R-4x. Two experimental tests, R-3x and R-4x, were constructed(4), each consisting of 15 pattern analysis, 10 arithmetic reasoning, and 10 reading and vocabulary items from the residue of those employed in constructing AGCT-3a, b, c, and d. Items were selected on the basis of high internal consistency and appropriate difficulty distribution. Items were paired for the two forms to give equivalence in terms of content, difficulty, and internal consistency. Directions for spatial items were expanded and illustrated more thoroughly in order to make the spatial items more understandable to lower level recruits.

Administration of Experimental Tests. Four populations were tested for this study from enlistees and inductees entering Camp Atterbury Reception Center, Massachusetts, between 15 April and 31 May 1946. Tests were administered to these populations as follows:

> Population A - R-2, AGCT-3a
> Population B - R-3x, AGCT-3a
> Population C - R-4x, AGCT-3a
> Population D - R-3x, R-4x, AGCT-3a

After testing larger groups, each population was chosen so as to duplicate proportionally the distribution of the general Army population in 1944 on AGCT, grade, and color. Examinees were asked to indicate the item on R-3x and R-4x reached at the end of 15 and 20 minutes, so that these time limits could be compared with the standard 25 minutes for those tests and the 15 minutes for R-2. The scoring formula was rights minus one-third wrongs for all tests.

Results of Administration of Experimental Tests. Preliminary results of this study showed a superiority of R-3x and R-4x over R-2 in terms of correlation with AGCT-3a at various points of cut, and of the 25-minute time limit for R-3x and R-4x over the 15- and 20-minute limits.

Appendix B shows N's, means, standard deviations, and correlations between the experimental tests and AGCT-3a. Both R-3x and R-4x correlated approximately .85 with AGCT-3a as compared to .79 between R-2 and AGCT-3a. The correlation between R-3x and R-4x was .80. These findings suggest that R-3x, R-4x, and AGCT-3a are homogeneous in content, and that the shorter tests R-3x and R-4x are somewhat less reliable as measures of this content.

Conclusions. 1. The 25-minute time limit for R-3x and R-4x gave better prediction of AGCT-3a within the desired range of scores than the 15- and 20-minute limits.

2. The R-3x and R-4x tests were more accurate predictors of AGCT-3a within the desired range than R-2.

3. The reliability of R-3x and R-4x was adequate.

4. The mean item difficulty (mean p-value) of R-3x and R-4x was somewhat low for optimal predictive efficiency within the range of primary interest.

Recommendations. It was recommended that R-3x and R-4x be introduced without change in item content and that revisions be initiated to decrease the mean item difficulty (i.e., to increase the mean p-values).

All three tests, R-2, R-3, and R-4, were standardized by determining raw score equivalents to AGCT-3a standard scores using the equipercentile method. Appendix C shows the standard score conversion tables derived for these three tests from the aforementioned Camp Atterbury group.

In August 1946, R-3 and R-4 were introduced as screening tests at Central Examining Stations, where a standard score of 70 or over qualified a man for enlistment. The R-2 test was transferred for use together with R-1 at local recruiting stations for pre-screening applicants before transmitting them to Central Examining Stations.

DEVELOPMENT OF R-5 AND R-6

Early in 1948, the responsibilities of Central Examining Stations for final examining of recruits was transferred to the Recruiting Service at Main Recruiting Stations. At this time it was decided that R-3 and R-4 would be used for pre-screening at local recruiting stations, replacing R-1 and R-2.

Another test was needed for final screening at Main Recruiting Stations. The AGCT-1c and AGCT-1d were republished as R-5 and R-6 and authorized for this purpose. While R-5 and R-6 carried booklet covers with the title, "Classification Test R-5" (or R-6), there were no changes in the content of the test. Therefore the norms previously developed in standardization of AGCT-1c and AGCT-1d were used for R-5 and R-6 without further standardization. Appendix D shows the norm table for converting R-5 and R-6 raw scores to Army Standard Scores.

Later in 1948 these same R-5 and R-6 tests were republished with new covers and entitled General Classification Test (GCT) -5 and -6.

DEVELOPMENT OF THE ARMED FORCES QUALIFICATION TEST (AFQT), FORMS 1 AND 2[1]/

TRANSITION FROM R-TESTS TO AFQT

The tests R-5 and R-6 (or GCT-5 and GCT-6) continued in use at Main Recruiting Stations until 1 January 1950, when they were replaced by AFQT-1 and AFQT-2. The development of screening tests R-1 through R-5 and R-6, had been a single Service (Army) endeavor. With the passage of the Selective Service Act of 1948, which provided that the Services would not reject anyone for mental reasons who had Army standard scores of 70 or better, the need arose for greater uniformity in mental screening procedures among the Services.

At this time three different tests were in use among the various branches of the Armed Forces to evaluate generally similar characteristics of inductees or applicants for enlistment. They were:

1/ Dr. H. Brandt was largely responsible for directing the development of AFQT-1 and -2, and for the first draft of this section of the report.

Army and Air Force: General Classification Test, Forms 5 and 6.
Navy: Navy Applicant Qualification Test, Form 3.
Marine Corps: Army General Classification Tests, Form 1c and 1d.

In addition, three different systems of converting scores and reporting results of tests were in use by the various Armed Forces, as follows:

Army and Marine Corps: Standard scores based on a mean of 100 and standard deviation of 20.

Navy: Standard scores based on a mean of 50 and standard deviation of 10.

Air Force: Standard scores based on a mean of 5 and a standard deviation of 2.

Anticipating a request for uniform screening instruments, an unofficial working group of technicians from the three Services began planning a joint screening test in 1948. On 26 November 1948, the Advisory Committee on Selective Service, Office of the Secretary of Defense, recommended that this working group be given official sanction as a subcommittee to study uniform screening tests and scoring systems for inductees and enlistees in the Armed Forces. The subcommittee[2] was officially so designated on 27 January 1949. It was by joint research that the Army, Navy, and Air Force, through this subcommittee, developed AFQT-1 and AFQT-2. The Department of the Army was given responsibility for direction of the project and for analysis and presentation of data.

SPECIFIC OBJECTIVES

The problems involved in the development of this common screening instrument were manifold. However, it was most fundamental to determine what the test should measure. It was decided that the test should represent a global measure of mental ability, containing essentially those item types which were most common to existing screening tests in all Services. Therefore, it was agreed that, as previous research experience indicated, the test would include items of the vocabulary, arithmetic reasoning, and spatial relations types.

Other problems of a more specific character soon became apparent.

1. The new instrument should have maximal sensitivity in the 60 to 90 Army Standard Score range as well as adequate distribution of scores throughout the total range so that division into five grade groups for allocation of personnel to the various Services could be accomplished.

2/ Subcommittee Military personnel were as follows: Navy--Cmdr. C. E. McCombs, Chairman; Army--Lt. Col. C. G. Dunn and Lt. Col. D. B. Routh; Air Force--Maj. Albert L. Klinge; Marine Corps--Lt. Col. B. D. Godbold; and Coast Guard--Cmdr. E. T. Callahan. Psychological Specialists appointed by this subcommittee as project personnel included: Dr. J. E. Uhlaner, D/A, TAGO, Program Coordinator; Dr. H. Brandt, D/A, TAGO, Project Director; Dr. E. G. Brundage, D/N, Bu Pers; Dr. J. H. Criswell, D/N, Bu Pers; and Dr. Frank A. Geldard, D/AF, HRD. Other technical specialists who were utilized in the development of AFQT-1 and -2 included: Dr. Glenn Finch, D/AF, HRD; Dr. Donald Baier, D/A, TAGO; Dr. C. I. Mosier, D/A, TAGO. The following sampling specialists advised on the design of the standardization plan: Dr. W. E. Deming, Div. Stat. Standards, Bureau of Budget; Dr. B. Tapping, Sampling Research Section, Bureau of Census; and Dr. D. Chapman, RDB.

2. The instrument would have to be used for both enlistment and induction, hence the populations to be used in the validation runs had to be considered in terms of a potential military population under emergency conditions rather than populations enrolled in the various Services during peacetime.

3. A uniform scoring system had to be developed, requiring either a completely new system or integration of the variations in the current score systems. The use of a percentile system was agreed upon.

4. The relationship of the new screening instrument to previous recruitment tests and to certain current classification tests had to be determined.

5. The relationship of the new instrument to background variables, such as age, sex, race, education, etc., had to be determined.

6. The character of the new instrument as a "power" test or a rigorous time limit test had to be decided. The decision was made to emphasize "power" rather than speed so as to avoid penalizing men with adequate mental ability but low motivation.

ANALYSIS OF TEST ITEMS

As was pointed out earlier, the decision was made to include vocabulary, arithmetic reasoning, and spatial relations items in the qualification test. Experience dictated also that the items be related to everyday activities in the Services and avoid the extreme academic and abstruse; that speed be minimized; that a difficulty range be obtained by fine discriminations in content, and that the verbal directions be simplified.

The vocabulary items were of two types:

1. Logical association--in which each choice is an aspect of the lead (e.g., a baby tries to manipulate anything he sees--choose one: chew, feel, handle, touch).

2. Pattern with similar affective tone--in which all the choices are similar with respect to positive or negative emotional tone (e.g., he is ill--choose one: hurt, pale, sick, sad).

In the arithmetic reasoning items the attempt was made to keep verbal and computational components at a minimum so as to emphasize the reasoning aspect of the items. The verbal element was simplified by using words rated as most frequently used according to the Thorndike-Lorge counts. Computation was reduced by using the ordinary range of number combinations and common fractions to such an extent that most items could be solved mentally.

Seven types of arithmetic reasoning items were employed in the test. They were: fundamental processes (whole numbers), number concept (indication of process, not computation), estimation (selection of closest rather than exact answer), fractions (all processes), ratio (proportion), percentage (all cases), and mensuration (use of various systems of weights and measures).

In the spatial relations items, a wide range of material was covered by using two and three dimensions. Several series of varying types of spatial items were constructed which embodied identification of simple objects (concrete and abstract), folding and unfolding patterns and forms (solids and cut-outs), and construction of wholes from parts or parts from wholes.

Collection of Item Analysis Data. A total of 554 items of the three foregoing types was selected and item indices (difficulty, validity, internal consistency, and independence) were determined after experimental administration of the tests to approximately 7,000 men entering the Army, Air Force, and Navy.

For the purpose of item analysis, the items for each type were divided equally into two separate booklets and arranged in order of guessed difficulty. These experimental booklets were labeled GCT-7x and GCT-8x. The tests were administered during the month of October 1948 to Army, Air Force, and Navy personnel at recruiting stations and training divisions, selected to give a good spread of the population on the reference tests. The experimental tests were administered following the official recruiting tests which were used to establish acceptance or rejection. The reference tests for the Army and Air Force populations were the recruiting tests GCT-5 and GCT-6[3]/ or AGCT-3a. For the Navy population the reference test was the recruiting test, Applicant Qualification Test (AQT-3).

A total of 7,114 cases was tested on both forms of GCT-7x and GCT-8x of which 389 were discarded because of lack of complete data. The breakdown by form number, by recruiting stations, and training divisions is shown in Appendix E.

Further adjustments were made to obtain the same racial proportions reported for the entire period of World War II. After these adjustments which reduced the number of Negroes to 10% and other non-whites to 1% of the total sample, 5,742 cases remained for the various phases of the item analysis.

For the purpose of item-difficulty analysis, all available cases (5,742) were used. Three populations were set up, one for each of the reference tests:

1. For the first (Navy) population, the cases obtained in the recruiting stations and training divisions were combined and AQT-3 was used as the reference test.

2. The second population was a combined Army-Air Force group from recruiting stations and training divisions, and the recruiting test, R-5 and R-6 was used as one part of the criterion.

3. The third population was drawn only from the training divisions of the Army and Air Force, and AGCT-3a was used as one part of the criterion.

These populations were normalized before item validities (biserial correlation coefficients) were obtained. For the two Army-Air Force populations a Mean of 100 and a Standard Deviation of 20 were used; for the Navy population the Mean was 50 and the Standard Deviation was 10.

For the purpose of obtaining item validities, three populations were set up based on each of the three reference tests. These populations were normalized before the biserial coefficients were obtained.

After the populations were normalized, their sizes were as shown in Table 2.

3/ Since GCT-5 and GCT-6 were the previous R-5 and R-6 tests, the latter designation will be used below for convenience in referring to the reference tests for the Army and Air Force populations.

Table 2. Populations used in item validity analysis of GCT-7x and GCT-8x.

Group	GCT-7x	GCT-8x
AQT-3 (Navy)	264	300
R-5, R-6 (Army-Air Force)	500	300
AGCT-3a (Army-Air Force)	400	558

For the internal consistency and independence analyses, it was not necessary to keep the populations distinct as to Service since scores for all testees on all parts of the tests were available. The three normalized populations used in the validity analysis were combined and split into two equally distributed halves (sample S1 and sample S2) shown in Table 3.

Table 3. Populations used in internal consistency and independence analyses of GCT-7x and GCT-8x.

Group	GCT-7x	GCT-8x
S1 - Split half of combined AQT-3, R-5, R-6, and AGCT-3a	582	579
S2 - Other half of combined AQT-3, R-5, R-6, and AGCT-3a	582	579

Sample S1 was used to obtain the biserial coefficients of each item with the total score of each of the three subtests. Sample S2 was used to obtain a second biserial coefficient of the item with the total score of the test of which it was a part.

Results of Item Analysis

1. Item difficulty analysis.

A new type of item difficulty index was developed in this study. The item difficulty level assigned was the lowest reference test score interval at which an item was answered correctly by at least 50% of the group in the interval. This was done for each of the two reference scores. The criterion scores were grouped into nine intervals and coded as shown in Table 4.

For each item in the experimental tests GCT-7x and GCT-8x, three difficulty indices were secured based on AQT-3, R-5 or R-6, and AGCT-3a scores.

Table 4. Equivalence of difficulty levels to standard score intervals on Navy and Army-Air Force tests.

Difficulty Level	Navy (AQT-3)	Army-Air Force (R-5, R-6, AGCT-3a)
1	29 and below	59 and below
2	30-34	60-69
3	35-39	70-79
4	40-44	80-89
5	45-49	90-99
6	50-54	100-109
7	55-59	110-119
8	60-64	120-129
9	65 and above	130 and above

2. Item validities.

The scores on the 112 vocabulary items of GCT-7x and GCT-8x were correlated with the items on the reference tests AQT-3, R-5, and R-6, and AGCT-3a. It can be seen from Table 5 that each of the reference tests yielded approximately the same average item validity for both experimental tests.

Table 5. Average item validities of the vocabulary items of GCT-7x and GCT-8x.

	GCT-7x			GCT-8x		
	AQT-3	R-5	AGCT-3a	AQT-3	R-6	AGCT-3a
Mean (item biserials)	.44	.45	.43	.45	.50	.45
Standard Deviation	.18	.15	.13	.18	.18	.12
Number of items	112	112	112	112	112	112

For the arithmetic reasoning items the average item validity (see Table 6) for the AQT-3 (Navy) criterion is slightly lower than for the other two on both experimental tests. This is understandable since the AQT-3 is more highly verbal than either the R-5, R-6, or AGCT-3a.

Table 6. Average item validitie: of the arithmetic reasoning items of GCT-7x and GCT-8x.

	GCT-7x			GCT-8x		
	AQT-3	R-5	AGCT-3a	AQT-3	R-6	AGCT-3a
Mean (item biserials)	.39	.47	.45	.44	.53	.52
Standard Deviation	.15	.10	.13	.16	.15	.14
Number of items	75	75	75	75	75	75

For the spatial relations items (see Table 7), the validities were about the same for the two reference tests used. The third test, AQT-3, was not used in this analysis because it contained no spatial items.

Table 7. Average item validities of the spatial relations items of GCT-7x and GCT-8x.

	GCT-7x		GCT-8x	
	R-5	AGCT-3a	R-6	AGCT-3a
Mean (item biserials)	.26	.30	.35	.31
Standard Deviation	.13	.13	.15	.12
Number of items	91	91	91	91

3. Internal consistency.

The mean item biserials for the original items were computed for each of the three types of content and for each sample (S1 and S2). In each case the particular total test score was the criterion. The values shown in Table 8 were obtained.

- 16 -

Table 8. Internal consistency coefficients for the items of GCT-7x and GCT-8x.

	Vocabulary 7x S1 S2	Vocabulary 8x S1 S2	Arithmetic Reasoning 7x S1 S2	Arithmetic Reasoning 8x S1 S2	Spatial Relations 7x S1 S2	Spatial Relations 8x S1 S2
Mean (item biserials)	.55 .64	.64 .61	.66 .62	.69 .58	.49 .49	.43 .48
Standard Deviation	.17 .25	.25 .24	.14 .11	.16 .20	.22 .15	.13 .13
Number of items	112 112	112 112	75 75	75 75	91 91	91 91

It can be seen that the average correlations are quite consistent for the two samples and are about the same for both experimental tests.

4. Independence.

To arrive at an index of independence, performance on an item of one type was correlated against the total score achieved on each of the other two types of item. For example, vocabulary performance was correlated against total score on arithmetic reasoning and spatial relations. The mean item biserials are given in Table 9.

Table 9. Independence coefficients for the items of GCT-7x and GCT-8x.

	Vocabulary Items — Arithmetic Reasoning Score 7x 8x	Vocabulary Items — Spatial Relation Score 7x 8x	Arithmetic Reasoning Items — Vocabulary Score 7x 8x	Arithmetic Reasoning Items — Spatial Relation Score 7x 8x	Spatial Relations Items — Vocabulary Score 7x 8x	Spatial Relations Items — Arithmetic Reasoning Score 7x 8x
Mean (item biserials)	.53 .56	.43 .27	.38 .47	.41 .35	.22 .31	.34 .09
Standard Deviation	.21 .21	.21 .11	.11 .12	.10 .13	.10 .10	.15 .13
Number of items	112 112	112 112	75 75	75 75	91 91	91 91

Examination of these coefficients reveals that performance on the spatial relations items is least related to performance on both the vocabulary and arithmetic reasoning items. The relatively high relationship between vocabulary and arithmetic reasoning is probably the result of the verbal element still present in the arithmetic items.

Selection of Items for AFQT, Forms 1 and 2. After the item analysis of GCT-7x and GCT-8x the task was to develop two forms of the screening instrument which came to be known as the Armed Forces Qualification Test (AFQT-1 and AFQT-2).

It was decided that the test should be short enough so that when fitted into 40 minutes of working time it would be a power rather than a speed test. Data obtained from the experimental administrations indicated that the final forms of the tests would best measure power if there were 90 items; 30 each for vocabulary, arithmetic reasoning, and spatial relations. Furthermore, the selection of the items was determined by the number of items attempted by 85% of the populations to which the experimental forms were administered.

Items were selected for the AFQT first on difficulty level with priority given to those items whose levels were identical for the three criterion groups. For those items whose difficulty levels were identical for only two of the criterion groups, that particular level was used. Where the difficulty levels of an item differed among the three criterion groups and the item had to be used, an average difficulty level was derived.

Each group of 30 items was distributed among the various difficulty levels as shown in Table 10.

Table 10. Distribution of difficulty level in AFQT, Forms 1 and 2.

Difficulty Level	1 and 2	3	4	5	6	7	8	9
GCT (Standard Score)	Below 70	70-79	80-89	90-99	100-09	110-19	120-29	130+
% (of total item content)	20	$16\frac{2}{3}$	$13\frac{1}{3}$	$16\frac{2}{3}$	10	10	$6\frac{2}{3}$	$6\frac{2}{3}$
No. of items	6	5	4	5	3	3	2	2

Once the items were sorted according to the nine difficulty levels, the magnitude of the over-all index became the next guide for selection. The over-all index was a weighted composite of item biserial correlation coefficients equal to the sum of the three validity coefficients plus the sum of the two internal consistency coefficients minus the sum of the two independence coefficients. This over-all index was an appropriate criterion of the general effectiveness of an item. In addition, the individual coefficients were carefully checked to insure the selection of items with the highest validity and internal consistency together with maximum independence.

The foregoing indices were considered also in regard to the construction of the two alternate forms of AFQT. In this connection, a "percent passing" index was obtained for the normalized sample and items were paired with respect to their p-values. In addition to matching by difficulty level, comparability of AFQT-1 and AFQT-2 was further assured by checking the items for similarity in content and psychological process.

In Table 11 are presented the results of pairing the items on difficulty level and the distributions for each of the three types of material for the two forms.

Table 11. Comparability of item difficulty levels in AFQT-1 and AFQT-2.

% Passing	Vocabulary AFQT-1	AFQT-2	Arith. Reasoning AFQT-1	AFQT-2	Spatial Relations AFQT-1	AFQT-2	Total AFQT-1	AFQT-2
95	3	3	-	-	-	-	3	3
90	5	2	2	3	-	-	7	5
85	1	3	2	4	1	1	4	8
80	-	2	6	2	1	2	7	6
75	3	3	1	4	4	1	8	8
70	2	2	2	2	2	4	6	8
65	3	2	3	3	3	4	9	9
60	1	-	2	1	3	4	6	5
55	2	4	1	1	3	2	6	7
50	4	-	1	1	4	3	9	4
45	1	3	2	-	1	2	4	5
40	1	2	1	3	2	1	4	6
35	-	1	2	-	2	2	4	3
30	2	1	1	1	2	1	5	3
25	-	-	-	1	1	1	1	2
20	1	1	1	2	1	-	3	3
15	1	1	2	1	-	1	3	3
10	1	-	1	1	-	1	1	2
N	30	30	30	30	30	30	90	90
Mean	66.83	66.17	61.00	62.83	57.33	57.17	61.20	61.55
Stand. Dev.	22.87	22.65	23.92	24.73	17.10	18.62	21.85	22.45

A listing of the item analysis results for the 90 items chosen for each form of AFQT is to be found in Appendix F. This Appendix contains the three difficulty indices, the percent passing, the three validity coefficients, the two internal consistency coefficients, the two independence coefficients, and the over-all index.

STANDARDIZATION OF AFQT

Population and Collection of Data. The population on which norms for the AFQT were established was a sample of the total potential military population under emergency mobilization conditions. The choice of this population was made because successful military operations require planning for mobilization and because it would aid in arriving at an equitable distribution of the available manpower pool among the various Services in the event of an emergency. The problem, then, was to decide what kind of sampling was needed to give a representation of the total potential military population. Two alternatives were possible. The entire civilian population could be sampled or use could be made of a previous population for which data were already available. The decision was made to use a previous population for the following reasons:

1. It was assumed that the millions of men available for testing prior to December 1944 would not differ essentially in age, education, occupational status, geographic distribution, etc., from a similar population to be utilized five or ten years later.

2. The use of data on hand would be more economical than testing a sample of the entire civilian population.

The population selected was that of all the men on duty in all the Services as of 31 December 1944. This included enlisted men, officer candidates, officers risen from the ranks, and officers who had been commissioned directly from civilian life. Since many of the officers who had been directly commissioned had not been tested, corrections were applied to the score distributions. These corrections proved to be minor (for further discussion, see Appendix G). The AFQT scores were assigned in accordance with the percentages falling between 110 and 162 based upon the obtained distribution for enlisted men.

All scores were converted to a common base, the Army Standard Score System. After the scores were converted, the distributions were blown up to the total 31 December 1944 strength (11,694,229). A composite cumulative percentile curve was set up in 5-point Army Standard Score intervals and the percentage of the total distribution was calculated for each interval. Appendix G presents a more detailed account of the derivation of this distribution of Army Standard Scores.

Four samples of 1,000 cases each were selected to reproduce the distribution for the entire population. AFQT-1 was administered to two of these samples. In one of the samples, AFQT-1 was administered before the reference test (order I); in the other, AFQT-1 was administered after the reference test (order II). AFQT-2 was administered to the other two samples in the same two orders. The purpose of the two orders was to control the effect of order of administration. Both orders for each form were used in the development of the conversion tables.

The Army's portion of the standardization population was obtained from three Army training divisions (3rd Armored at Fort Knox, Kentucky, 5th Armored at Camp Chaffee, Arkansas, and 10th Infantry at Fort Riley, Kansas). The Air Force's portion came from the Air Force Indoctrination Center at Lackland AFB, Texas, and the Navy's portion was obtained from training centers at Great Lakes, Illinois, and San Diego, California. The cases consisted of all incoming new recruits at these installations. They were tested during July 1949. Additional testing was necessary in order to fill in the gaps at both ends of the population which were caused in part by the use of enlistment cutting scores by the three Services. These additional cases were obtained by the Army from selected installations.

Statistical Treatment. The AFQT's for each of the four groups were scored and plots were made of the cumulative percentile distributions of raw scores for each order of the two forms of the test. The scoring formula used was rights only. An examination of these percentile curves showed very little discrepancy in the two orders of administration for either form. Accordingly, the orders were combined and the distributions for each form were plotted. The differences between the two forms, particularly at the proposed cutting points, were so slight that the use of a single conversion table appeared justified.

The similarity of the two forms of the test was further confirmed by another study in which the two forms were compared on an additional Army group of 600 cases. For this group there was an average practice gain of only two raw score points on either form. A correlation of .93 between the two forms was obtained. As a result of this similarity between the two forms, the distributions of the four standardization samples (4,000 cases) were combined into a single distribution and a single percentile curve was plotted.

By means of equipercentile conversion, the AFQT scores were translated into Army Standard Scores. Thus, any AFQT raw score or equivalent percentile score could be interpreted in terms of the conventional Army scale.

Results. The percentiles obtained were compared with the expected percentiles of the normal curve. This comparison for the standard score more commonly used administratively is shown in Table 12.

Table 12. Comparison of expected and obtained percentiles for AFQT standard scores.

Standard Score	Percentiles Expected	Percentiles Obtained
130	93.3	93
120	84.1	82
110	69.2	65
100	50.0	49
90	30.9	31
80	15.9	21
70	6.7	13
60	2.3	7
50	0.6	3

It can be seen from Table 12 that the obtained percentiles were fairly close to the expected for standard scores above 80. Below the standard score of 80, there was a greater discrepancy.

One interpretation of this finding is that at the low end of the distribution, the score achieved is influenced by both lack of mental ability and by illiteracy, as suggested by the content of the test. In line with this interpretation, it is advisable to supplement the verbal type of tests with non-verbal materials which will permit illiterates to demonstrate their mental ability, provided that special training is made available for illiterates.

Appendix H shows conversions of AFQT-1 and AFQT-2 raw scores to percentiles and Army Standard Scores as derived from this standardization study.

The relationship of AFQT with previous Army aptitude tests was examined. AFQT was found to be highly correlated with the reference test AGCT-1c or AGCT-1d, regardless of the order of administration (Table 13).

Table 13. Correlations between AGCT and AFQT.

	AGCT-1c or AGCT-1d with:	AFQT-1		AFQT-2		AFQT Total Group
		(order I)	(order II)	(order I)	(order II)	
Correlation (r)		.91	.90	.91	.90	.90
Mean	99.6	54.9	55.9	56.7	56.9	56.1
Standard Deviation	22.7	19.0	18.7	19.3	18.9	19.0
N	4000	1000	1000	1000	1000	4000

- 21 -

AFQT was found to be substantially correlated with the individual tests comprising Aptitude Area I (Reading and Vocabulary, Arithmetic Reasoning, and Pattern Analysis), and even higher with Aptitude Area I (Table 14). These correlations were obtained for the Army portion of the AFQT-1 administered in the order I sample only.

Table 14. Correlation between AFQT and Aptitude Area I tests.

	AFQT with:	Reading and Vocabulary	Arithmetic Reasoning	Pattern Analysis	Aptitude Area I
Correlation (r)		.83	.87	.75	.92
Mean	53.5	97.7	89.4	95.7	94.3
Standard Deviation	18.3	23.5	24.5	25.8	21.9
N	552	552	552	552	552

The number of years of education was correlated with scores on AFQT (order I) and R-5 and R-6 (Table 15).

Table 15. Correlations between years of education and AFQT-1, R-5, and R-6. (N = 929)

	Mean	Standard Deviation	Correlation With Years of Education
AFQT-1	55.0	19.3	.69
R-5, R-6	98.9	22.8	.67
Years of Education	10.2	2.1	

AFQT depends less on speed than does R-5 and R-6. Distributions of the last item attempted showed that 51% of the men completed AFQT-1, 64% completed AFQT-2, and only 4% completed the R-tests. Another comparison showed that on the average, 92% of the items on AFQT-1 were attempted, 94% on AFQT-2, and only 64% on R-tests.

OPERATIONAL APPLICATION OF AFQT-1 AND AFQT-2

Enlistment. AFQT-1 and AFQT-2 were put in operation for screening of enlistees at Navy and joint Army-Air Force recruiting stations on 1 January 1950. The two forms of AFQT replaced the R-5 and R-6 tests for screening at Main Recruiting Stations and local

recruiting stations continued to use R-3 and R-4 for pre-screening before forwarding applicants to Main Stations.   The AFQT then became the common mental screening instrument for all Armed Forces.

Induction.   Induction had been used very little by the Army during the months at the end of 1948 and beginning of 1949. However, in July 1950, under the 1950 extension of the Selective Service Act, the Army began active procurement of inductees through Selective Service.   The Department of the Army was designated as executive agent for Joint Armed Forces Examining and Induction Stations which would process inductees for all Armed Forces at such time as the other Services should call upon Selective Service for procurement of personnel. Regulations for screening inductees provided for use of AFQT-1 and AFQT-2 with a minimum acceptable score of percentile 13 (raw score 31).   These regulations also provided that, should other Services place calls for inductees, registrants would be allocated to the Services proportionally within mental groups I, II, III and IV. Mental group V is composed of those scoring below the cut-off score.

Introduction of the "Converted Score." As AFQT-1 and AFQT-2 were used in induction and recruiting, it was noted that the norms for AFQT in terms of Army Standard Score equivalents did not seem to be the same for operationally obtained results as those established in the original AFQT standardization. This was particularly noted at reception centers where the Army Classification Battery was being administered. An abnormally large proportion of new men who had passed AFQT at induction and recruiting stations were found to score below the equivalent Army Standard Score on Aptitude Area I of the Army Classification Battery.   Aptitude Area I was composed of components of the old AGCT-3a and AGCT-3b and had been previously found to correlate .92 with AFQT. This pronounced discrepancy between AFQT and Aptitude Area I scores was termed "Operational Slippage," and was attributed to non-standard administration of AFQT at induction and recruiting stations.

In order to correct conversions of AFQT for non-standard administration of the test in the field, a new conversion table of raw scores into "Converted Scores" was placed into effect 10 July 1950, for all Services using Selective Service. This table is shown in Appendix I. The new table did not explain what a "Converted Score" was. However, in appearance it is similar to a percentile score. Its range is from 1 to 100, and it takes the form of an ogive curve when raw scores are plotted against "Converted Scores." In the new table there was considerable agreement between percentile and "Converted Scores" equivalents to AFQT raw scores in the upper quarter of the range, but up to 17 points difference in the lower half and middle of the range. This table of "Converted Scores" replaced the percentile norms for AFQT until 1 December 1951, when the original percentile norms for AFQT-1 and AFQT-2 were restored.   At this time it was expected that with the assignment of personnel officers to supervise the testing at Armed Forces Examining Stations, standard testing conditions would be maintained.

Allocations.   On 1 May 1951, by direction of the Secretary of Defense the policy of qualitative division of military manpower accessions among the Services on an equitable basis was established.   This policy applied to the total of male enlistee and inductee accessions with certain exceptions such as officer candidates, aviation cadets, and prior service enlistees.   This policy provided that the number of enlistees and inductees procured by each Service must conform to a fixed percentage distribution among the first four AFQT mental groups. The percentages of input for each Service were allocated as shown in Table 16.

It should be mentioned that the "Converted Score" equivalents to mental group limits were the same numerically as the Percentile Score limits for these mental groups which had

- 23 -

Table 16. Percentage allocation by mental group of enlisted and inducted manpower among the Services.

Mental Group	AFQT "Converted Score"	AFQT Raw Score	Manpower Percentage Allocated
I	93 - 100	82 - 90	8
II	65 - 92	71 - 81	32
III	31 - 64	57 - 70	39
IV	13 - 30	39 - 56	21

been established for inductions under SR 615-180-1, 27 April 1950. Consequently, Army Standard Score equivalents for the mental group limits were at considerable variance with the traditional uniform pattern.

With the passage of the Universal Military Training and Service Act, which lowered the minimum AFQT acceptance score for military service, the lower limit of mental group IV was changed to "Converted Score" 10 and allocation quotas were adjusted in accordance with a Department of Defense directive as shown in Table 17.

Table 17. Percentage reallocation of enlisted and inducted manpower among the Services.

Mental Group	AFQT "Converted Score"	AFQT Raw Score	Manpower Percentage Allocated
I	93 - 100	82 - 90	8
II	65 - 92	71 - 81	31
III	31 - 64	57 - 70	38
IV	10 - 30	34 - 56	23

These definitions of mental group limits and the allocation percentage quotas were further adjusted following the studies discussed in the next Section of this report.

FOLLOW-UP STUDY OF AFQT-1 AND AFQT-2 STANDARDIZATION

PURPOSE

Since there was no controlled study which preceded the establishment of the "Converted Score" norms placed into effect for AFQT-1 and AFQT-2, questions were raised regarding some of the facts of the situation. The primary questions were: (1) To what extent was there "Operational Slippage" in the originally established Aptitude Area I standard score equivalents to AFQT scores; and (2) What is the nature of the distribution of operationally administered AFQT scores and of its relation to that obtained under controlled administration?

In order to clarify the facts underlying the original standardization of AFQT and the norms as shown in the "Converted Score" conversion table, the Assistant Chief of Staff, G-1 directed that a study be undertaken to compare and evaluate the test scores obtained at recruiting and induction stations in relation to scores on the same tests administered under more uniform conditions of administration and motivation and in relation to Aptitude Area I scores obtained in initial processing; so as to determine whether any change in existing AFQT score conversions is indicated, and, if so, what change should be made.

GENERAL PROCEDURE

The study was designed so that procedures, duplicating the original AFQT standardization, would be carried out separately for: (1) AFQT scores obtained from induction and recruiting station administration; and (2) Scores on the alternate form of AFQT for the same men, obtained at training divisions under standardized conditions. Standardization was accomplished by the equipercentile method, using Aptitude Area I as a reference test.

POPULATION

The sample consisted of 1 000 men undergoing reception processing at Fort Dix, New Jersey; Fort Knox, Kentucky; Fort Jackson, South Carolina; and Fort Riley, Kansas, during the week of 12 February 1951. This sample was selected (from a total of 4,981 men constituting the regular flow of input at that time) to duplicate proportionally, by 10-point standard score intervals in R-5, the World War II population. It included both enlistees and inductees, unselected as such, but drawn as shown in Table 18.

Table 18. Population components used in study of AFQT "Converted Scores."

	Fort Dix	Fort Knox	Fort Jackson	Fort Riley	Total
Enlistees (No.)	50	82	97	59	288
Inductees (No.)	199	174	158	181	712
Total (No.)	249	256	255	240	1000

RESULTS

Comparison of Standardization Runs. Results of the original AFQT standardization were compared with the follow-up standardization results for AFQT using scores from both induction and recruiting stations (called "operational conditions") and from standard administration at the training divisions (called "standard conditions"). Table 19 shows these results in terms of AFQT raw score equivalents (computed by the equipercentile method) for certain Army Standard Scores.

Table 19. Comparative AFQT-1 and AFQT-2 norms from standard and operational administrations.

Army Standard Score	AFQT Raw Score Equivalents		
	Original Standardization (Standard Cond.)* (1)	Present Study (Standard Cond.)** (2)	Present Study (Operational Cond.)** (3)
160	90	90	90
150	88	90	90
140	85	89	89
130	81	84	83
120	74	77	75
110	65	68	67
100	57	57	57
90	48	49	47
80	39	41	43
70	31	31	39
60	22	20	36

*Standard score equivalence determined using R-5 (AFQT) as a reference test.
**Standard score equivalents determined using Aptitude Area I as a reference test.

It can be seen that this study gave approximately the same norms for AFQT given under standard conditions as were given in the original standardization study.

Comparison of AFQT Scores Obtained from Operational and from Standard Test Administration. However, it may also be seen in Table 19 that there were more pronounced differences in AFQT standard score equivalents between the original standardization study and results obtained from AFQT tests given under operating conditions. The AFQT equivalents to standard scores of 80 and below were considerably higher for operational scores than they were in the original standardization. Particularly, a standard score of 70 increased from AFQT raw score 31 to raw score 39; and a standard score of 60 increased from raw score 22 to raw score 36. Standard score conversions for AFQT, then, did not hold up at levels below standard score 80 for AFQT scores obtained under operational administration conditions.

The most dramatic evidence of such inflation at the lower end of the distribution was shown in the comparisons of raw score distributions for AFQT administered under operational and under standard conditions (see Figure 1). The distribution of scores from standard administration is relatively smooth, though negatively skewed. This is a typical unselected sample distribution for AFQT. The skewness has been built into the test, since it was constructed to be more differentiating at lower portions of the scale. The distribution for operationally obtained AFQT scores is markedly bimodal, and most abnormally modal at the interval containing raw score 39, the current Army cut-off score. The absence of scores below 39, of course, is due to applications of this cut-off, plus the fact that administratively accepted inductees were not included in this study.

" Administratively accepted inductees" refers to men who were inducted because their failure on AFQT was interpreted to be motivational rather than genuine lack of ability. High school graduates who failed were administratively determined by the commanding officer of the examining station to have met the mental requirements. Other failures were interviewed and administrative determination was made. A Terminal Screening Guide was prepared at a later date to serve as an aid to the interviewer by providing suggestions as to the type of additional information he might find useful in arriving at his decision (e.g., job history, educational history, ability to drive a car, instructions for administering the Individual Examination).

The obvious interpretation of these graphs is that most of the scores around 39 which are obtained at induction and recruiting stations represent scores for men whose scores in AFQT obtained under standardized conditions really are lower.

Scatterplots of operational administration and standardized administration AFQT scores showed this to be true. The distribution of standardized administration scores of those in the interval containing operationally administered raw score 39 predominantly ranged below raw score 39.

The above mentioned inaccuracy in operational AFQT scores at the lower end of the distribution occurs in scores reported from both induction and recruiting stations. Figures J-1 and J-2 in Appendix J show the comparative distributions of standard administration scores and operational administration scores for both enlistees and inductees. Again the abnormal mode was obtained at the operational score interval containing the cut-off score of 39 for each group. In the enlistee operational distribution 29% of scores were in this interval as were 16.5% of the inductee scores, showing that the abnormality was more pronounced for enlistees. Separate scatterplots of operational and standard administration AFQT scores for enlistees and inductees also demonstrated that operational scores reported at and immediately above the cut-off score 39 are predominantly for both enlistees and inductees whose standard administration scores were at failure levels.

CONCLUSIONS

The major conclusions based on the results of the comparison of AFQT scores obtained under operational conditions and under standard conditions were:

1. "Operational Slippage" occurred in the operational scores as indicated by the pile-up of scores at the cut-point, for both enlistees and inductees.

2. Under standard administration, the pile-up at the cut-point was absent and the original standardization percentiles were obtained.

3. The correction of operating conditions of test administration appears to be the solution for avoiding the discrepancies which appear in operational AFQT scores.
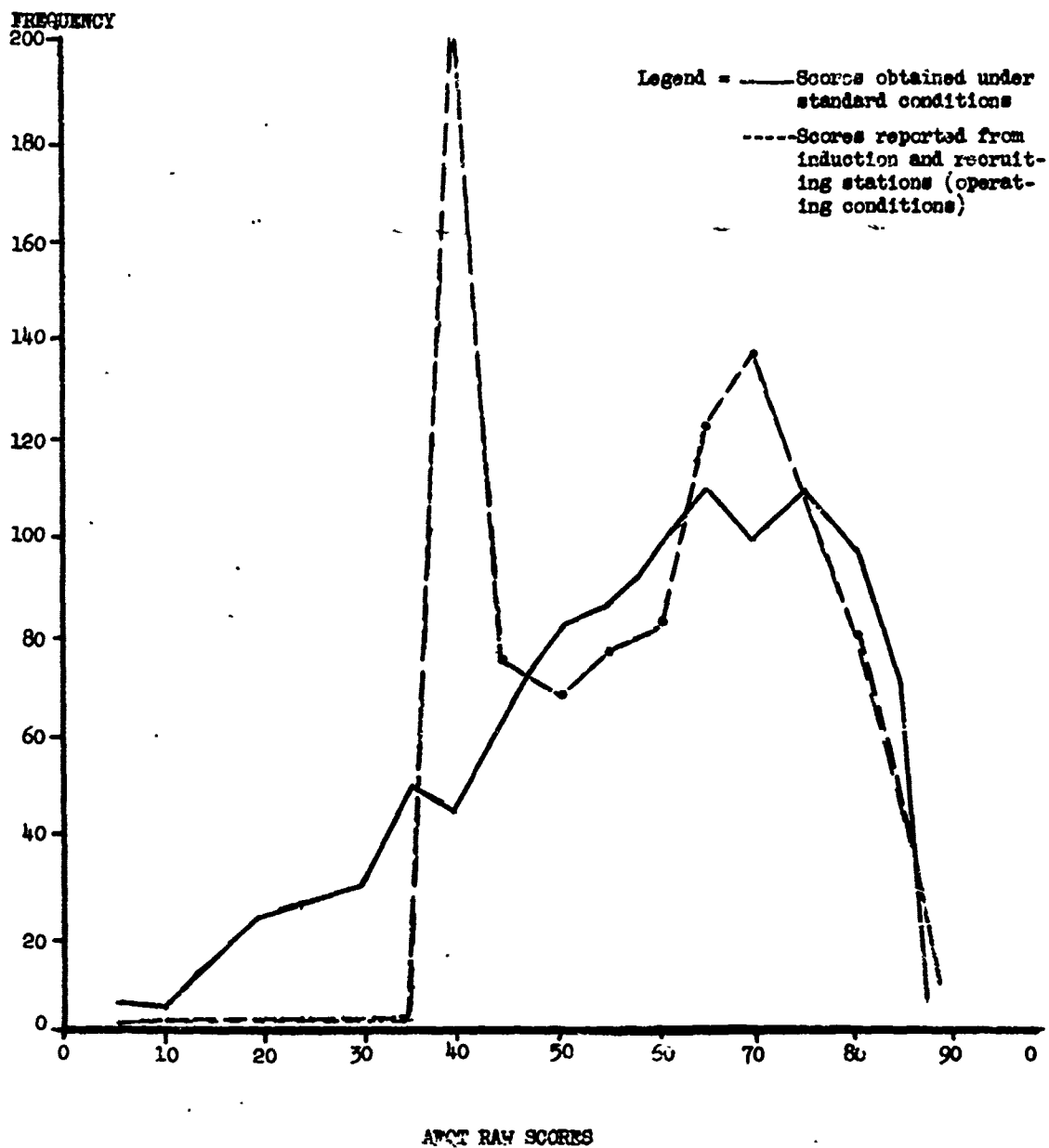
Figure 1. AFQT scores of same 1,000 men from operational and standard administrations.

EVALUATION OF AFQT CONVERSION TABLES

Along with the expansion of military strength, there was an increasing interest on the part of the Department of Defense in manpower problems. This was evidenced by a strengthening of the organization of the Office of the Assistant Secretary of Defense, Manpower and Personnel, to provide direct channels for dealing with inter-service manpower questions. In the area of recruitment and induction such a channel was provided by the organization of a system of Armed Forces Examining Stations in 1951 to deal with questions of procurement, selection, and allocation of military manpower. Beginning 1 July 1951, the examining functions of Army, Navy, Air Force, and Marine Corps Main Recruiting Stations were consolidated and responsibility for these transferred to Armed Forces Examining Stations. By the end of 1951 there were about 75 such stations throughout the United States. Responsibility for development of policy applicable to these stations was vested with the "Armed Forces Examining Station Policy Board" (AFES PB). This Board was established within the Office of the Assistant Secretary of Defense, Manpower and Personnel. It was composed of the Director, Manpower Utilization, Office of the Assitant Secretary of Defense, Manpower and Personnel, as Chairman, and one general or flag officer from each of the four Services. This Board designated the Department of the Army as executive agent for administration of the various Armed Forces Examining Stations, though staffing of the stations included personnel from all Services.

One of the primary problems of the AFES Policy Board at this time was that of allocation of military accessions (recruits and inductees) to the four Services. As was pointed out in the previous discussion on operational application of AFQT, such accessions were allocated on the basis of distributions among the mental groups I, II, III, and IV as determined by AFQT. This allocation began on 1 May 1951. However, considerable concern arose immediately over the fact that in experience of the Services during May and June 1951, the obtained percentage distributions of accessions for the four mental groups differed appreciably from the predicted percentages as prescribed in the allocation formula. Table 20 shows the predicted (prescribed) percentages for allocation and the obtained percentages of all Services' input in the four AFQT mental groups.

Table 20. Percentages of accessions in AFQT Mental Groups--1951.

Mental Group (1)	"Converted Score" Range (2)	Prescribed Quotas 2 April Directive (3)	Actual Distribution	
			May (4)	June (5)
I	93-100	8.0	5.5	6.4
II	65-92	32.0	16.8	17.9
III	31-64	39.0	27.6	30.8
IV	13-30	21.0	50.1	44.9
Total	13-100	100.0	100.0	100.0

The prescribed quotas shown in column (3) were deemed to be an adequate prediction of the distribution of total input, based on the assumption that the "Converted Scores" represent percentile norms. Therefore, any mental group should contain the proportion of total input

equal to the range of "Converted Scores" in the group divided by the total range of acceptable scores (i.e., range in mental group II equals 65-92 or 28; total acceptable range equals 13-100 or 88; therefore expected percentage in group II equals 28/88 or 32 as shown in column (3)). When the obtained distribution for May and June did not resemble the prescribed quotas, the AFES Policy Board raised the question of the accuracy of the AFQT norm table.

To answer this question the AFES Policy Board appointed a "Working Group on Evaluation of the AFQT Conversion Table" to evaluate the correctness of the norms. This working group consisted of one research psychologist from each of the four Services.[4]

This working group recognized the differences between the norms for AFQT as originally established in the original percentile norms and as established in the revised "Converted Score" norms. It also took cognizance of the fact that the "Converted Score," although it represents a partial correction for non-standard operational test administration, does not define percentile norms for AFQT. Therefore, using the original standardization percentile equivalents to AFQT "Converted Scores," it was shown that a predicted distribution similar to that obtained in May and June would result.

The working group recommended that:

1. Test administration practices in Armed Forces Examining Stations be improved to attain standard conditions.

2. When such conditions are attained, the original percentile norms be used to replace "Converted Score" norms in determining allocation quotas.

3. Based on these percentile scores under the current standards for induction (minimum acceptable score at the 10th percentile), the quotas for the four mental grades should be as shown in Table 21.

Table 21. Expected percentages of accession under standard testing conditions.

Mental Grade	Percentage Quotas
I	9
II	31
III	37
IV	23
Total	100

---

[4] The members were: Army--Dr. Julius E. Uhlaner; Navy--Dr. Kenneth E. Clark; Marine Corps--Mr. Francis F. Medland; Air Force--Dr. Charles C. Limbury, Chairman.

Operational action was taken almost immediately to effect improvement of conditions highlighted by these studies. Two major actions had a direct bearing on the above recommendations:

The administration of mental testing was placed under the supervision of commissioned personnel psychologists assigned to the AFES. The Army set up a special training program at the Adjutant General's School in which personnel psychologists (MOS 2230) were given a two-week intensive course specific to AFES mental examining procedures. Emphasis was placed in this course on the necessity for good test administration to control "Operational Slippage" in AFQT. Each AFES was subsequently assigned at least one of these trained personnel psychologists. Preliminary evidence points in the direction of considerable improvement.

The AFES Policy Board re-instituted the original AFQT percentile conversion table 1 December 1951, to replace the "Converted Score" table.

USE OF AFQT BY THE ARMY, NAVY, MARINE CORPS, AND AIR FORCE

The Armed Forces Qualification Test is currently used as an initial screening instrument by all four military Services. The various uses to which the test is put by the Services are summarized below. All cutting scores were administratively determined.

ALL SERVICES

1. The AFQT is used for determination of acceptability for enlistment on the basis of mental qualifications. The minimum acceptable score on AFQT-1 or AFQT-2, beginning 15 July 1951, is percentile score 10.

2. The AFQT is used for determination of acceptability for induction of Selective Service registrants on the basis of mental qualifications whenever any Service uses the induction machinery for procurement of personnel. The minimum acceptable score on AFQT-1 and AFQT-2, beginning 15 July 1951, is percentile score 10.

3. In induction screening, an additional use of AFQT is made for classifying those registrants who fail to achieve the minimum score. For this purpose, the answer sheets of AFQT failures are rescored with the AFQT Verbal-Arithmetic Key, which provides a rights minus one-third wrongs score on the first 36 verbal and arithmetic items. Those failures achieving such a score of 6 or higher are designated as not acceptable, but are placed in a deferred category for possible future induction.

4. Equitable distribution of military accessions (allocation) is accomplished by percentage quotas of total chargeable accessions within each Service based on distribution of men in AFQT mental groups I through IV.

ARMY

1. The AFQT is used for determining acceptability of women for enlistment in the WAC on the basis of mental qualifications. The minimum AFQT acceptable score is percentile score 31.

2. Applicants for enlistment from civilian life as officer candidates are screened with AFQT to determine mental qualification. The minimum acceptable score on AFQT is percentile score 65.

NAVY

Mental qualifications of women enlistee applicants for the WAVE is determined by the AFQT. The minimum acceptable score is percentile score 37.

MARINE CORPS

Women applicants for enlistment in the Marine Corps are screened with AFQT. The minimum acceptable score is percentile score 37.

AIR FORCE

Women applicants for enlistment in the WAF are screened with AFQT. The minimum acceptable score is percentile score 49.

SUMMARY OF PROGRAM

This program was inaugurated to develop screening tests which would provide a basis for mental standards of acceptance of recruits and inductees.

1. In response to the Army's need for screening tests (as distinguished from classification tests), R-2, R-3, and R-4 were constructed. These tests were first used at Central Examining Stations for final screening. Following the development of R-5 and R-6 and their introduction operationally at Central Examining Stations, R-3 and R-4 were available for use by local recruiting stations as initial screening tests.

2. To meet the need for greater uniformity among the Services in mental screening procedures, The Armed Forces Qualification Tests (AFQT), Forms 1 and 2 were developed as a joint effort of Army, Navy, Air Force, and Marine Corps research personnel, with the Department of the Army acting as coordinating agency. The AFQT is used operationally by the Army, Navy, Air Force, and Marine Corps for determining mental qualifications of male and female enlistee applicants and for screening Selective Service registrants for induction by those Services which may so procure personnel. The AFQT also provides the basis for the Department of Defense system of qualitative distribution of military accessions among the four Services.

3. Each of the two forms of AFQT contains 90 items divided equally among vocabulary, arithmetic reasoning, and spatial relations. Items were selected on the basis of item analyses so as to provide a spread of difficulty over the entire useful range. To provide comparable forms, items were matched not only for validity and difficulty but for similarity in content and psychological process as well.

4. Standardization of AFQT was based on samples of the entire military population on duty in all the Services as of 31 December 1944. The two forms were standardized separately. The differences in the distribution of scores on the two forms were so slight that a single conversion table was adopted. By means of equipercentile conversion, scores were translated into standard forms.

5. AFQT scores were found to be highly correlated with scores on Army aptitude tests such as AGCT and its successor Aptitude Area I.

6. It was found that distributions of scores obtained from operational administration differed significantly from the distribution of scores expected on the basis of the standardization studies. A follow-up study substantiated the original standardization. To reduce this "Operational Slippage" steps were taken to control test administration at Armed Forces Examining Stations. Preliminary evidence points in the direction of considerable improvement.

7. Other studies have been and still are being made. One series of studies has resulted in the construction of "motivation keys" to control the effect of attempts to distort or bias test scores. Another series is directed at developing nonverbal forms of mental screening tests. Research efforts will continue to be directed toward maintaining and improving screening tests in accordance with improvements in test construction techniques, administrative policies, and operating problems.

COLLECTION OF DATA: February 1946 to November 1951

PREPARATION OF REPORT: 1 May 1952

REFERENCES

Reports of the Personnel Research Section, Personnel Research and Procedures Branch,
The Adjutant General's Office, Department of the Army.

1. PRS Report 69. Classification test R-1. 22 April 1941.

2. PRS Report 271. Selection of items from GCT-1b for R-2. 28 February 1942.

3. PRS Report 309. Standardization of R-2. 25 April 1942.

4. PRS Report 718. The development and evaluation of classification tests R-3 and R-4.
   14 June 1946.

5. PRS Report 778. Comparison of Army and Navy classification tests. 29 April 1949.

APPENDIXES

APPENDIX A

CHRONOLOGY OF SCREENING TESTS AND STANDARDS
(Tests R-1 through AFQT-1 and AFQT-2)

18 April 41: Regulations excluded from military service all men who did not have the capacity for "reading and writing the English language as commonly prescribed for the fourth grade in grammar school." (Changes to Mobilization Regulation 1-7, 18 April 1941)

1 June 42: Induction was permitted of men who could not meet the literacy standards provided they possessed "sufficient intelligence to absorb military training rapidly." (WD Circular 166, 1 June 1942)

31 Oct 42: R-1 introduced as screening test for induction of limited service (physically restricted) personnel. Minimum acceptable score is standard score 90. (TWX, Symbols OC-S-WDGAPO, 31 Oct 1942)

Nov 45: Inductions stopped. Procurement by enlistment only begun.

7 Feb 46: G-1 directed that a test be constructed for use in Army enlistment which could be administered individually or in groups with a minimum of verbal instruction, could be completed in less than 30 minutes, and would provide scores as nearly equivalent to AGCT-3a standard scores as possible within the lower test score ranges. (D/F from WDGS, G-1, File: WDGAP 342.06, 7 February 1946)

17 April 46: Raw score of 15 on R-1 (standard score 70) is minimum acceptable for enlistment. (WD Circular 110, 17 April 1946)

12 June 46: R-1 used at local recruiting stations, where raw score of 15 is minimum acceptable for referral to Central Examining Station, where raw score 16 on R-2 (standard score 70) is minimum acceptable for enlistment. R-2 transferred from Main Recruiting Stations to Central Examining Stations. (WD Circular 171, 12 June 1946)

9 Aug 46: R-1 and R-2 used at local recruiting stations where raw scores of 15 on R-1 or 16 on R-2 (standard score 70 on either) are minimum acceptable for referral to Central Examining Stations, where raw score of 6 on R-3 or R-4 (standard score 70) is minimum acceptable for enlistment. (WD Circular 239, 9 August 1946)

23 April 47: Raw score of 19 on R-1 or 23 on R-2 (standard score 80 on either) are minimum acceptable at local recruiting stations for referral to Central Examining Stations, where raw score of 13 on R-3 or R-4 (standard score 80) is minimum acceptable for enlistment. R-2 is to replace R-1 as soon as stocks of R-1 are depleted. (WD Circular 103, 23 April 1947)

11 Feb 48: Examining functions are transferred from Central Examining Stations to Main Recruiting Stations. Use of R-5 and R-6 is prescribed. (Memo 600-750-28, 11 February 1948)

12 March 48: Minimum acceptable score for enlistment at Main Recruiting Stations is standard score 80 on R-3, R-4, R-5, or R-6. (WD Circular 66, 12 March 1948)

30 March 48: Standards are the same as in WD Circular 66. Tests R-3 or R-4 is given at local recruiting stations and R-5 or R-6 is given at Main Recruiting Stations. (Memo 600-750-30, 30 March 1948)

27 July 48: Minimum score acceptable for enlistment is standard score 70 on R-3 or R-4, but 80 on R-5 and R-6. (Memo 600-750-30, 27 July 1948)

Nov 48: Inductions began under Selective Service Act of 1948. Inductions continued for 3-month period, (November, 1948--January, 1949) and were then terminated until August 1950. Minimum acceptance standards on R-5 and R-6 same as for enlistment (standard score 70 GCT equivalent included in Selective Service Act of 24 June 1948, PL 759-80th Congress-as the minimum acceptable score).

30 Dec 48: Minimum acceptable score at initial recruiting point is raw score 9 on R-3 or R-4 (standard score 75) for referral to Main Recruiting Stations (where standard score 70 on R-5 or R-6 is minimum acceptable for enlistment). (WCL 39061, 30 December 1948)

1 Jan 50: AFQT Forms 1 and 2 to replace R-5 and R-6 at Main Recruiting Stations. At local recruiting stations R-3 and R-4 continue in use. (SR 615-105-25, 13 December 1949)

1 Jan 50: R-3 and R-4 used at local recruiting stations where raw score of 17 or higher on either (standard score 90) is minimum acceptable for referral to Main Recruiting Stations where a percentile score of 31 (standard score 90) on AFQT-1 or AFQT-2 is minimum acceptable for enlistment. (TAG letter, File: AGSE 342, 22 December 1949)

27 April 50: AFQT Forms 1 and 2 to be used at Joint Examining and Induction Stations for screening <u>inductees</u>. Percentile score 13 (standard score 70) to be minimum acceptable for induction. (SR 615-180-1, 27 April 1950)

10 July 50: A new conversion table of raw scores into "Converted Scores" was prepared at the direction of Army G-1 and placed into operation 10 July 1950. (Letter from TAG to all Army Commands, File: AGPP-P 220.01, 10 July 1950)

17 July 50: For enlistments, raw score 6 on R-3 (standard score 70) is minimum acceptable at local recruiting stations for referral to Main Recruiting Stations where percentile score of 13 (standard score 70) is minimum acceptable for enlistment. (WCL 32938, TAG, 17 July 50)

18 July 50: For enlistment, standards at local recruiting stations unchanged. At Main Recruiting Stations, "Converted Score" 13 on AFQT-1 or AFQT-2 (standard score 70 adjusted for operational administration) is minimum acceptable for enlistment. (WCL 33372, TAG, 18 July 1950) "Converted Score" table replaced percentile score equivalent table for determining AFQT-1 and AFQT-2 norms to adjust for slippage in operational testing conditions for enlistment.

Aug 50: Inductions begin for Army under 30 June 1950 Extension of Selective Service Act of 1948. (PL 599-81st Congress)

2 Nov 50: "Converted Score" 13 is minimum acceptable score on AFQT-1 and AFQT-2 (adjusted standard score 70) for <u>induction</u>. (SR 615-180-1, Change 3, 2 November 1950)

11 Jan 51: G-1 directed that a study be undertaken to compare test scores obtained under operational and under standard conditions so as to check the existing AFQT score conversions. (DF from G-1 to TAG, File: G-1 201.6, 11 January 1951)

2 April 51: By direction of the Secretary of Defense, the policy of qualitative division of military manpower accessions among the Services on an equitable basis was established. (Memorandum for Secretaries of Army, Navy, and Air Force, and Joint Chiefs of Staff, Subject: "Qualitative Distribution of Military Manpower," 2 April 1951)

19 June 51: Percentile score of 10 (standard score 65) or AFQT established as minimum acceptable for induction. (PL 51, 82nd Congress, amendment to Universal Military Training and Service Act)

30 June 51: "Converted Score" of 10 on AFQT-1 and AFQT-2 (adjusted standard score 65) is minimum acceptable for induction. (Department of Defense Directive 100.03-1, 30 June 51)

18 July 51: "Converted Score" of 10 on AFQT-1 and AFQT-2 (adjusted standard score 65) is minimum acceptable for enlistment at Main Recruiting Stations. (DA Radio 34878, TAG, 18 July 1951)

30 Oct 51: Examining functions for recruitment and induction by all Services transferred to Armed Forces Examining Stations. Commissioned personnel psychologists were assigned to supervise administration of testing. (SR 615-100-1, 30 October 1951)

5 Nov 51: For induction, AFQT-1 and AFQT-2 supplemented by additional screening with the AFQT Verbal-Arithmetic Subtest, Non-Language Qualification Test (NQT-1). This supplemental screening given AFQT failures to classify them for possible future induction. (SR 615-180-1, 5 November 1951)

23 Nov 51: Percentile score of 10 on AFQT-1 and AFQT-2 (standard score 65) is minimum acceptable for both enlistment and induction. The "Converted Score" table for determining AFQT norms is replaced by the original percentile norm table, since standard testing conditions are assumed to have been achieved in examining stations. (DA Radio 46247, TAG, 23 November 1951)

N's, MEANS, STANDARD DEVIATIONS AND PRODUCT-MOMENT CORRELATIONS
FOR INDICATED COMBINATIONS OF R-2, R-3, R-4, AND AGCT-3a

Camp Atterbury Reception Center, May 1946

Form	Population*	N	Rx-Test		AGCT-3a		r Between Rx-Test and AGCT-3a
			M	SD	M	SD	
R-2	A	700	34.4	13.2	98.8	20.5	.79
R-3x	B	700	22.1	10.9	98.9	20.6	.86
R-3x	D	1000	22.5	11.4	98.8	20.1	.83
R-4x	C	600	22.1	10.7	99.0	20.6	.86
R-4x	D	1000	25.6	11.3	98.8	20.1	.86
							r Between R-3x and R-4x .80

*Tests administered: A - R-2, AGCT-3a
                      B - R-3x, AGCT-3a
                      C - R-4x, AGCT-3a
                      D - R-3x, R-4x, AGCT-3a

CONVERSION TABLES: RAW SCORES ON R-2, R-3, AND R-4 TO ARMY STANDARD SCORES

Raw Scores R-2	Raw Scores R-3 and R-4	Standard Scores	Raw Scores R-2	Raw Scores R-3 and R-4	Standard Scores
	45	144		22	99
			38		98
	44	142	37	21	97
	43	138	36	20	96
			35		95
50	42	135	34	19	94
			33		93
	41	132	32	18	92
			31		91
49	40	130	30	17	90
			29		88
	39	128	28	16	87
			27		85
	38	126		15	84
					83
48	37	125	26		83
			25	14	82
			24		81
	36	123	23	13	80
	35	122		12	79
	34	121	22	11	78
			21		77
47	33	120	20	10	76
	32	119	19 and 18	9	75
			17		74
	31	117	16	8	73
46	30	115	15 and 14	7	72
	29	113	13 and 12	6	71
45		112	11		70
	28	111	10	5	69
			9		68
44	27	109	8	4	66
			7		64
43	26	107		3	62
			6		61
			5		59
42	25	106		2	58
			4		57
			3		55
41	24	104	2	1	53
40		102	1		51
	23	101	0	0	47
39		100			

CONVERSION TABLE: RAW SCORES ON R-5 AND R-6 TO ARMY STANDARD SCORES

Raw Score	Standard Score	Raw Score	Standard Score	Raw Score	Standard Score	Raw Score	Standard Score
140	163	100	139	65	110	25	74
139	162	99	138	64	110	24	73
138	162	98	137	63	109	23	72
137	161	97	136	62	109	22	71
136	161	96	135	61	108	21	70
135	160	95	134	60	108	20	69
134	160	94	133	59	107	19	68
133	159	93	132	58	106	18	67
132	159	92	131	57	105	17	66
131	158	91	130	56	104	16	65
130	158	90	129	55	103	15	64
129	157	89	129	54	102	14	63
128	157	88	128	53	101	13	62
127	156	87	128	52	100	12	61
126	156	86	127	51	99	11	60
125	155	85	127	50	99	10	59
124	155	84	126	49	98	9	58
123	154	83	125	48	97	8	57
122	154	82	124	47	96	7	56
121	153	81	124	46	95	6	54
120	153	80	123	45	94	5	52
119	152	79	122	44	93	4	50
118	152	78	121	43	92	3	48
117	151	77	120	42	91	2	46
116	151	76	120	41	90	1	44
115	150	75	119	40	89		
114	150	74	119	39	89		
113	149	73	118	38	88		
112	149	72	117	37	87		
111	148	71	116	36	86		
110	147	70	115	35	85		
109	146	69	114	34	83		
108	145	68	114	33	82		
107	144	67	113	32	81		
106	143	66	112	31	80		
105	142			30	79		
104	141			29	78		
103	140			28	77		
102	140			27	76		
101	139			26	75		

BREAKDOWN OF TEST POPULATION FOR GCT-7x AND GCT-8x BY DIVISION, STATION, AND STATUS

	GCT-7x			GCT-8x		
	Rejects	Recruits	Total	Rejects	Recruits	Total
Recruiting Stations						
Army-Air Force	158	430	588	149	456	605
Navy	129	---	129	134	---	134
Total	278	430	717	263	456	739
Training Divisions						
Army (Dix)			1125			1015
Air Force (Lackland)			772			715
Navy (San Diego)			861			781
Total			2758			2511
Combined Recruiting Stations and Training Divisions						
Army			1713			1620
Air Force			772			715
Navy			990			915
Total			3475			3250

Table F-1. Basic item statistics of items in final form of AFQT-1.

AFQT Item No.	Experimental Form	Experimental Item No.	Difficulty Level AQT	Difficulty Level AOQT	Difficulty R	% Right	Validity (Biserials) AQT	Validity (Biserials) AOQT	Validity (Biserials) R	Consistency Item-Test Self Correlations Sample₁	Consistency Item-Test Self Correlations Sample₂	Independence Coefficient Vocabulary	Independence Coefficient Arithmetic Reasoning	Independence Coefficient Spatial Relations	Overall Index
1-8	(practice items)														
9	8X-V	56	1	1	1	95	.54	.62	.68	.91	.86	••	.76	.32	2.53
10	8X-V	44	1	1	2	96	.38	.47	.75	1.12	.97	••	.89	.34	2.46
11	8X-V	20	1	1	1	95	.46	.45	.54	.77	.92	••	.63	.15	2.37
12	8X-V	24	2	2	3	93	.78	.99	..	1.13	.99	••	.78	.20	3.18
13	8X-V	29	2	2	2	92	.58	.62	.65	1.03	.83	••	.76	.38	2.57
14	8X-V	40	2	2	2	91	.68	.50	.61	.91	1.04		.65	.33	2.76
15	7X-A	14	1	1	2	88	.29	.42	.64	.80	.71	.42	••	.45	.02
16	7X-A	9	1	1	2	92	.26	.26	.45	.81	.74	.54	••	.74	1.68
17	7X-A	12	1	1	1	87	.40	.26	.45	.49	.58	.20	••	.28	1.68
18	8X-A	35	2	3	3	85	.58	.62	.71	.81	.63	.69	••	.46	2.22
19	7X-A	27	2	3	3	83	.39	.48	.51	.65	.62	.42	••	.48	1.75
20	7X-A	23	3	1	4	78	.52	.41	.52	.68	.65	.39	••	.47	1.92
21	7X-B	6	1	1	1	88	••	.44	.40	.74	.59	.22	.49	••	1.45
22	7X-B	56	1	1	1	79	••	.42	.43	.70	.61	.32	.50	••	1.24
23	7X-B	10	1	1	3	79	••	.35	.37	.65	.57	.28	.44	••	1.22
24	7X-B	60	2	1	1	81	••	.33	.25	.67	.57	.21	.37	••	1.24
25	7X-B	26	3	2	2	79	••	.26	.29	.52	.55	.20	.42	••	1.00
26	7X-B	13	1	1	2	78	••	.27	.41	.56	.59	.19	.39	••	1.25
27	8X-V	55	1	1	3	91	.64	.50	.65	.95	.80	••	.74	.29	2.51
28	7X-V	52	3	3	3	80	.61	.66	.74	.89	.92	••	.79	.60	2.43
29	7X-V	51	3	3	4	80	.61	.61	.57	.76	.79	••	.59	.58	2.17
30	7X-V	83	3	2	4	79	.53	.48	.58	.73	.82	••	.54	.47	2.13
31	8X-V	77	4	4	4	79	.58	.59	.71	.89	.79	••	.57	.42	2.47
32	8X-V	73	4	4	4	78	.68	.66	.71	.89	.84	••	.72	.46	2.60
33	8X-A	12	2	2	3	84	.41	.36	.55	.56	.41	.48	••	.23	1.58
34	7X-A	16	1	3	2	81	.48	.51	.47	.67	.63	.47	••	.44	1.89
35	7X-A	29	3	4	3	83	.48	.48	.44	.71	.64	.42	••	.51	1.82
36	7X-A	15	2	3	4	81	.58	.66	.61	.81	.72	.54	••	.55	2.29
37	8X-A	38	2		3	76	.50	.42	.51	.66	.50	.53	••	.35	1.71
38	7X-A	38	3	5	3	71	.39	.51	.62	.72	.71	.43	••	.45	2.07
39	7X-B	58	2	2	3	73	••	.39	.34	.59	.55	.29	.33	••	1.20
40	7X-B	31	3	4	4	69	••	.38	.32	.53	.55	.23	.39	••	1.16
41	7X-B	39	1	4	4	70	••	.42	.33	.63	.62	.32	.37	••	1.31
42	8X-B	33	3	4	4	66	••	.30	.42	.41	.39	.35	.36	••	0.81
43	8X-B	40	3	2	3	63	••	.43	.47	.57	.59	.34	.39	••	1.33
44	7X-B	47	4	4	3	62	••	.40	.42	.62	.64	.34	.49	••	1.25
45	8X-V	35	3	4	3	73	.50	.57	.68	.73	.73	••	.60	.39	2.22
46	7X-V	39	4	4	4	73	.45	.52	.61	.65	.80	••	.52	.41	2.10
47	8X-V	76	4	4	4	69	.50	.52	.62	.70	.63	••	.58	.36	2.03
48	8X-V	83	5	5	5	68	.62	.54	.59	.81	.66	••	.64	.40	2.18
49	8X-V	84	6	6	5	66	.59	.63	.63	.85	.70	••	.65	.48	2.27
50	8X-V	80	6	5	5	63	.46	.46	.46	.71	.55	••	.33	.63	1.80
51	7X-A	34	5	4	5	70	.51	.41	.46	.60	.65	.33	••	.39	1.91
52	8X-A	41	4	5	5	69	.62	.61	.66	.78	.70	.64	••	.42	2.31
53	7X-A	43	4	5	4	68	.54	.59	.57	.78	.70	.49	••	.48	2.21
54	8X-A	39	3	4	4	66	.48	.55	.58	.71	.61	.54	••	.38	2.0.
55	7X-A	48	3	6	3	59	.59	.67	.60	.78	.82	.45	••	.53	2.48
56	7X-A	46	6	6	5	61	.54	.57	.56	.76	.75	.48	••	.44	2.26
57	7X-B	28	3	4	4	64	.54	.34	.35	.38	.48	.27	.23	••	0.95
58	8X-B	16	1	4	5		••	.39	.48	.41	.36	.23	.31	••	1.16
59	8X-B	51	6	6	5	29	••	.33	.48	.44	.50	.23	.32	••	1.14
60	8X-B	41	5	5	3	29	••	.39	.40	.60	.54	.33	.38	••	1.22
61	7X-B	25	5	6	4	34	••	.49	.32	.56	.59	.26	.43	••	1.30
62	7X-B	69	7	6	4	28	••	.21	.42	.51	.47	.21	.35	••	0.89
63	7X-V	71	6	6	6	55	.64	.58	.48	.56	.72	••	.53	.33	2.10
64	7X-V	84	6	6	6	55	.60	.54	.60	.72	.74	••	.50	.45	2.19
65	8X-V	91	7	5	6	53	.48	.49	.42	.69	.55	••	.50	.38	1.78
66	8X-A	43	6	6	5	61	.54	.52	.62	.66	.71	.45	••	.33	2.21
67	8X-A	42	6	6	6	48	.58	.58	.67	.78	.69	.57	••	.45	2.28
68	8X-A	26	5	5	5	50	.57	.71	.70	.87	.81	.60	••	.46	2.60
69	8X-B	58	5	5	5	58	••	.41	.52	.60	.53	.39	.43	••	1.24
70	7X-B	22	8	6	6	54	••	.31	.31	.22	.50	.30	.37	••	0.92
71	7X-V	71	6	6	6	50	••	.37	.44	.65	.58	.34	.45	••	1.26
72	7X-V	86	7	7	6	53	.56	.28	.57	.65	.68	••	.54	.41	2.03
73	8X-V	102	6	6	6	51	.54	.46	.48	.69	.63	••	.55	.34	1.91
74	7X-V	75	7	6	6	50	.39	.49	.29	.57	.75	••	.51	.44	1.85
75	8X-A	45	7	7	6	47	.66	.63	.62	.79	.78	.57	••	.42	2.48
76	8X-A	61	6	7	6	43	.66	.65	.65	.89	.82	.57	••	.53	2.58
77	8X-A	62	6	6	6	36	.58	.72	.64	.87	.79	.60	••	.55	2.45
78	7X-B	80	5	7	6	49	••	.36	.34	.63	.53	.34	.41	••	1.11
79	7X-B	83	7	6	6	43	••	.29	.22	.57	.52	.21	.27	••	1.12
80	7X-B	63	6	6	6	41	••	.38	.35	.54	.52	.24	.37	••	1.19
81	7X-V	112	6	6	6	49	.67	.29	.55	.69	.73	••	.58	.45	2.85
82	7X-V	97	8	7	7	41	.28	.51	.52	.64	.63	••	.53	.32	1.91
83	7X-V	98	9	8	7	34	.66	.58	.64	.66	.73	••	.57	.37	2.33
84	8X-A	58	7		7	38	.60	.67	.64	.82	.75	.56	••	.50	2.43
85	7X-A	63	7	7	7	34	.61	.65	.64	.85	.80	.59	••	.47	2.96
86	8X-A	63	8	8	8	35	.56	.70	.57	.81	.44	.50	••	.42	2.40
87	7X-B	29	8	7	6	38	••	.50	.50	.70	.67	.41	.55	••	1.41
88	7X-B	57	8	8	9	68	••	.50	.25	.32	.57	.05	.43	••	1.27
89	7X-B	20	7	7	7	31	••	.24	.27	.46	.48	.12	.22	••	1.00
90	8X-V	107	9	8	8	32	.54	.36	.56	.74	.69	••	.55	.34	2.20
91	8X-V	112	9	9	9	20	.60	.67	.59	.69	.6.	••	.59	.57	2.03
92	7X-V	107	9	9		18	.58	.38	.41	.58	.45	••	.48	.33	1.63
93	8X-A	70	9	9	8	19	.59	.36	.61	.70	.70	.42	••	.48	2.26
94	8X-A	73	8	8	8	16	.45	.61	.36	.69	.69	.36	••	.44	2.20
95	8X-A	74	9	8	8	13	.52	.53	.66	.80	.70	.48	••	.33	2.30
96	7X-B	18	8	8	8	36	••	.58	.49	.67	.72	.40	.54	••	1.52
97	7X-B	8.	9	9	9	20	••	.35	.09	.37	.40	.04	.17	••	1.02
98	8X-B	83	9	9	9	14	••	.37	.33	.56	.46	.31	.37	••	1.0.

1/ The level at which 50% of the criterion population passed the item
2/ An average based on 3 normalized samples with GCT mean of 100 and σ of 20, or Navy equivalent
3/ Based on normalized samples with GCT mean of 100 and σ of 20, or Navy equivalent
4/ Based on random sample from composite of 3 normalized samples used for validity coefficients
5/ 3 validity coefficients and 2 consistency coefficients - 2 Independence coefficients
• No computations made since AQT has no space component.
•• These correlations reported in self correlation (internal consistency) columns.

Table F-2. Basic item statistics of items in final form of AFQT-2.

AFQT Item No.	Experimental		Difficulty[1]		/[2]	Validity[3]			Consistency[4] Item-Test Self Correlations		Independence Coefficients[5]			Overall[6]
	Form	Item No.	AGF	AOU R	Right	AGF AOU R			Sample₁ Sample₂		Vocabulary	Arithmetic Reasoning	Spatial Relations	Index

APPENDIX G

DERIVATION OF DISTRIBUTION OF AGCT STANDARD SCORES FOR TOTAL STRENGTH POPULATION OF THE ARMED FORCES AS OF 31 DECEMBER 1944

The plan for standardization of AFQT-1 and AFQT-2 required that norms for the tests represent percentile scores in a total potential military population under conditions of full mobilization. In addition, since the Selective Service Act of 1948 fixed the minimum acceptable mental standard for induction as Army Standard Score 70, it was required that scores on the new AFQT equivalent to AGCT standard scores be established. This latter objective could be accomplished by determining equipercentile equivalent AFQT raw scores to AGCT standard scores for the sample population used in establishing the percentile norms.

The major sampling problem therefore was to obtain a standardization sample which represented the potential military population under conditions of full mobilization. It was agreed that the best model of such a population would be the total military population at the time of peak mobilization in World War II. This was in December 1944. It was further decided that a sample population which duplicated the AGCT standard score distribution of this designated parent population (World War II military strength) if generally controlled on geographical distribution and Service membership, would serve as a satisfactory sample for standardization. These agreements were based on two major assumptions: (1) The 11,694,229 enlisted men and officers in the Armed Forces as of December 1944 would not differ in significant population parameters from a potential full mobilization population in the next five or ten years; and (2) The distribution of AGCT standard scores would suffice as the major parameter on which samples could be selected to represent this population for standardization purposes. Table G-1 shows the total military manpower on 31 December 1944 as it was distributed among the various Services. It was necessary to break out numbers of my commissioned officers from those commissioned in schools, as shown, in order to AGCT standard score distributions from data which were available.

Table G-1. Strength of Armed Services as of 31 December 1944.

Service (1)	ENLISTED MEN (2)	OFFICERS Directly Commissioned (3)	Commissioned From Ranks (4)	Total Manpower (5)
Army-Air Force	7,127,897	220,543	619,940	7,968,380
Navy	2,735,270	293,268	82,716	3,111,254
Coast Guard	147,865	11,707	480	160,052
Marine Corps	414,561	11,995	27,987	454,543
TOTAL MANPOWER				11,694,229

The purpose of this Appendix is to demonstrate how the AGCT standard score distribution for this total World War II population was estimated from available data.

GENERAL PROCEDURE

The general procedure for estimating the AGCT score distribution for the total Armed Forces strength was: (1) To obtain an estimate of the distribution for each Service separately; and (2) To weight the distribution for each Service in accordance with its total strength (as shown in Table G-1) and to combine the weighted distributions.

INDIVIDUAL SERVICE DISTRIBUTIONS

As a basis for building AGCT distributions, very large samples were obtained of input in the various Services as follows:

Army-Air Force. A 2% sample of input during 1944 in 5-point AGCT standard score intervals. Since the Air Force was a part of the Army at this time, the two Services were treated as one.

Navy. Total input for the year January 1944 to February 1945 in terms of Navy General Classification Test scores. These were converted to equivalent AGCT scores by means of conversions available in a previous Army-Navy Classification Battery comparison study(5).

Marine Corps. Total input for the same period covered by the Navy sample. The Marine Corps used AGCT in its classification procedures, so no further conversions of scores were necessary.

It was assumed that the above distributions of enlisted men input would adequately represent the distribution of enlisted men and officers commissioned from the ranks as of the base period, December 1944, since this portion of the population had come from such input in the past. Therefore, the single correction of each distribution for officers directly commissioned from civilian life was necessary to give a representative distribution of strength as of December 1944.

CORRECTION FOR DIRECTLY COMMISSIONED OFFICERS

Data from Table G-1 gave the basis for applying corrections to each interval of each Service input distribution in order to account for directly commissioned officers. The following proportions of directly commissioned officers to enlisted men were derived from the table:

> Army Air Force: 7,127,897 enlisted men to 220,543 directly
> commissioned officers, or .969987 to .030013.
> Navy: 2,735,270 to 293,268, or .903165 to .096835.
> Marine Corps: 414,561 to 11,995, or .971879 to .028121.

Therefore, the input distribution of each Service was cast into a percentile distribution, which in turn was applied to the above proportion of enlisted men to show the proportional distribution of enlisted men. These distributions are shown in Table G-2, columns (2), (5), and (8). To these, the directly commissioned officers were added with their total proportion being distributed in intervals at standard score 110 and above in the same manner as was the distribution of enlisted men. The assumption here was that officers would be distributed like a random selection of enlisted men with standard scores 110 and above (assuming directly commissioned officers to be similar in quality to those selected from the ranks where AGCT standard score 110 is required for OCS).

- 48 -

Table G-2. Proportional distributions of AGCT standard scores for enlisted input and directly commissioned officers by Service.

AGCT Standard Score (1)	ARMY-AIR FORCE Enlisted Input (2)	Dir Com Off (3)	Total (4)	NAVY Enlisted Input (5)	Dir Com Off (6)	Total (7)	MARINES Enlisted Input (8)	Dir Com Off (9)	Total (10)
160 and up	.000026	.000003	.000029	.003352	.000935	.004287	.000096	.000007	.000103
155-159	.000362	.000035	.000397	.001777	.000496	.002273	.000391	.000023	.000314
150-154	.001940	.000187	.002127	.003941	.001100	.005041	.000973	.000080	.001053
145-149	.003880	.000375	.004255	.009349	.002609	.011958	.002042	.000169	.002211
140-144	.008730	.000843	.009573	.012483	.003483	.015966	.003985	.000331	.004316
135-139	.016490	.001592	.018082	.021353	.005958	.027311	.010400	.000865	.011265
130-134	.025414	.002454	.027868	.022880	.006385	.029265	.015841	.001315	.017156
125-129	.045589	.004402	.049991	.056789	.015847	.072636	.039552	.003294	.042846
120-124	.062855	.006070	.068925	.052073	.014531	.066604	.062395	.005186	.067581
115-119	.065959	.006370	.072329	.069877	.019499	.089376	.089801	.007454	.097265
110-114	.079539	.007682	.087221	.093147	.025992	.119139	.112932	.009387	.122319
105-109	.083225		.083225	.080819		.080819	.142089		.142089
100-104	.077793		.077793	.101135		.101135	.100493		.100493
95-99	.071973		.071973	.094966		.094966	.107297		.107297
90-94	.064019		.064019	.081581		.081581	.078721		.078721
85-89	.063049		.063049	.057872		.057872	.067351		.067351
80-84	.054319		.054319	.051286		.051286	.039458		.039458
75-79	.050439		.050439	.034311		.034311	.035571		.035571
70-74	.045783		.045783	.026291		.026291	.025462		.025462
65-69	.038799		.038799	.007109		.007109	.017397		.017397
60-64	.032980		.032980	.010147		.010147	.010885		.010885
55-59	.027160		.027160	.006772		.006772	.004567		.004567
50-54	.019400		.019400	.001983		.001983	.001554		.001554
45-49	.012319		.012319	.001262		.001262	.000973		.000973
40-44	.017942		.017942	.000610		.000610	.001653		.001653
TOTAL	.969987	.030012	1.000000	.903165	.096835	1.000000	.971879	.028121	1.000000

Table G-3. Proportional distribution of AGCT standard scores for total strength of Armed Forces as of 31 December 1944.

AGCT Standard Score (1)	Army-Air Force (2)	Navy (3)	Marines (4)	Total (5)	Cumulative (6)	Smoothed Percentiles (7)
160 and up	.000020	.001199	.000004	.001223	1.000000	.100
155-159	.000271	.000636	.000012	.000919	.998777	.100
150-154	.001449	.001410	.000041	.002900	.997858	.100
145-149	.002899	.003345	.000086	.006330	.994958	.100
140-144	.006523	.004466	.000168	.011157	.988628	.99
135-139	.012321	.007640	.000438	.020399	.977471	.98
130-134	.018989	.008186	.000667	.027842	.957072	.96
125-129	.034064	.020319	.001669	.056052	.929230	.92
120-124	.046965	.018632	.002627	.068224	.873178	.87
115-119	.049284	.025002	.003781	.078067	.804954	.80
110-114	.059432	.033328	.004754	.097514	.726887	.73
105-109	.056709	.022608	.005523	.084840	.629373	.63
100-104	.053008	.028291	.003906	.085205	.544533	.55
95-99	.049042	.026565	.004170	.079777	.459328	.47
90-94	.043622	.022821	.003060	.069503	.379551	.37
85-89	.042961	.016189	.002618	.061768	.310048	.30
80-84	.037013	.014347	.001534	.052894	.248280	.26
75-79	.034369	.009598	.001383	.045350	.195386	.20
70-74	.031196	.007355	.000990	.039541	.150036	15
65-69	.026437	.001989	.000676	.029102	.110495	12
60-64	.022472	.002838	.000423	.025733	.081398	.9
55-59	.018507	.001894	.000177	.020578	.055680	.6
50-54	.013219	.000555	.000060	.013834	.035082	.4
45-49	.008394	.000353	.000038	.008785	.021248	.2
40-44	.012228	.000171	.000064	.012463	.013463	.2
TOTAL	.681394	.279737	.058869	1.000000		

Table G-2, columns (3), (6) and (9), shows the corrections for directly commissioned officers. Columns (4), (7), and (10), show the total (sum of preceding two columns in each case) proportional distribution for each Service. These total proportional distributions were then taken to represent AGCT standard score distributions for total strength in each Service as of December 1944.

COMBINING ALL SERVICES

The individual percentage distributions for each Service were combined to yield a percentage distribution of AGCT standard scores for the total Armed Forces strength as shown in Table G-3. Each Service proportion of the total strength, 11,694,229 men, was derived by dividing this number into its total strength as shown in Table G-1. These proportions were as follows: Army-Air Force, .681394; Marine Corps, .038869; Navy, .279737. The Navy proportion included the Coast Guard strength shown in Table G-1 on the assumption that the Coast Guard personnel would be distributed in approximately the same manner as the Navy. For each Service, this proportional figure was multiplied by the total percentage in each AGCT interval as shown in Table G-2 to give the proportion of the total Armed Forces strength in each interval. The interval percentages for each Service were then added to yield the total Armed Forces distribution shown in column (5) of Table G-3. These were cumulated in column (6) to show the AGCT percentile norms for the World War II full mobilization population.

As a further extension the cumulative percentages were plotted and fitted to a smoothed ogive. Column (7) shows the final norms derived from the smoothed ogive. Subsequent samples used in standardizing AFQT were selected to duplicate this smoothed ogive curve.

APPENDIX B

CONVERSION TABLE: RAW SCORES ON AFQT-1 OR AFQT-2 TO PERCENTILE SCORES AND TO ARMY STANDARD SCORES

Raw Score	Percentile	Standard Score	Raw Score	Percentile	Standard Score
90	100	164	45	28	86
89	100	157	44	27	85
88	100	151	43	26	84
87	100	146	42	24	83
86	99	142	41	23	82
85	98	139	40	22	81
84	97	137	39	21	80
83	96	134	38	20	79
82	95	131	37	19	78
81	93	130	36	18	77
80	92	128	35	17	76
79	90	126	34	16	75
78	89	125	33	15	73
77	87	123	32	14	71
76	85	122	31	13	70
75	84	121	30	12	69
74	82	120	29	12	68
73	80	118	28	11	66
72	78	117	27	10	65
71	76	116	26	9	64
70	74	115	25	9	63
69	73	114	24	8	62
68	71	113	23	7	61
67	69	112	22	7	60
66	67	111	21	6	59
65	65	110	20	5	57
64	63	109	19	5	56
63	61	107	18	4	55
62	59	106	17	4	53
61	57	105	16	3	52
60	55	104	15	3	50
59	53	103	14	2	48
58	51	101	13	2	47
57	49	100	12	2	45
56	47	99	11	2	43
55	45	98	10	2	42
54	43	97	9	2	41
53	41	96	8	1	41
52	39	95	7	1	41
51	37	94	6	1	40
50	36	93	5	1	39
49	34	92	4	1	39
48	32	91	3	1	39
47		90	2	1	39
46	30	88	1	1	39

CONVERSION TABLE: RAW SCORE ON AFQT FORM 1 CR FORM 2 TO CONVERTED SCORE (10 JULY 1950)

| Raw Score	Converted Score	Raw Score	Converted Score	Raw Score	Converted Score
90	100	60	37	30	8
89	100	59	35	29	7
88	99	58	33	28	7
87	98	57	31	27	6
86	97	56	30	26	6
85	96	55	29	25	6
84	95	54	28	24	5
83	94	53	27	23	5
82	93	52	26	22	5
81	91	51	25	21	4
80	89	50	24	20	4
79	87	49	23	19	4
78	85	48	22	18	3
77	83	47	21	17	3
76	81	46	20	16	3
75	79	45	19	15	3
74	76	44	18	14	2
73	73	43	17	13	2
72	70	42	16	12	2
71	67	41	15	11	2
70	64	40	14	10	2
69	61	39	13	9	2
68	58	38	12	8	1
67	55	37	11	7	1
66	52	36	11	6	1
65	49	35	10	5	1
64	46	34	10	4	1
63	43	33	9	3	1
62	41	32	9	2	1
61	39	31	8	1	1
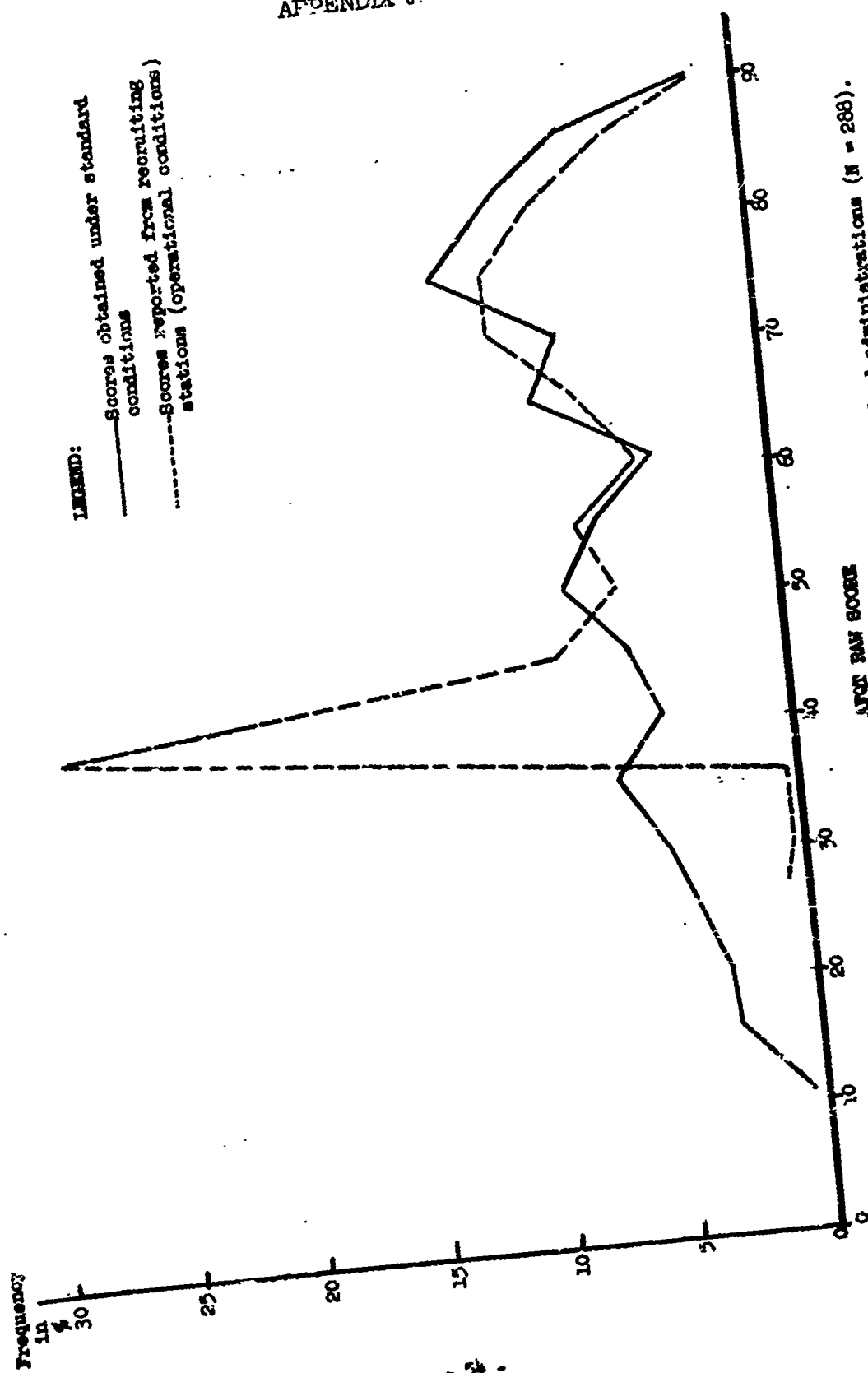
Figure J-1. AFQT scores of same enlistees from operational and standard administrations (N = 288).

AFQT RAW SCORE

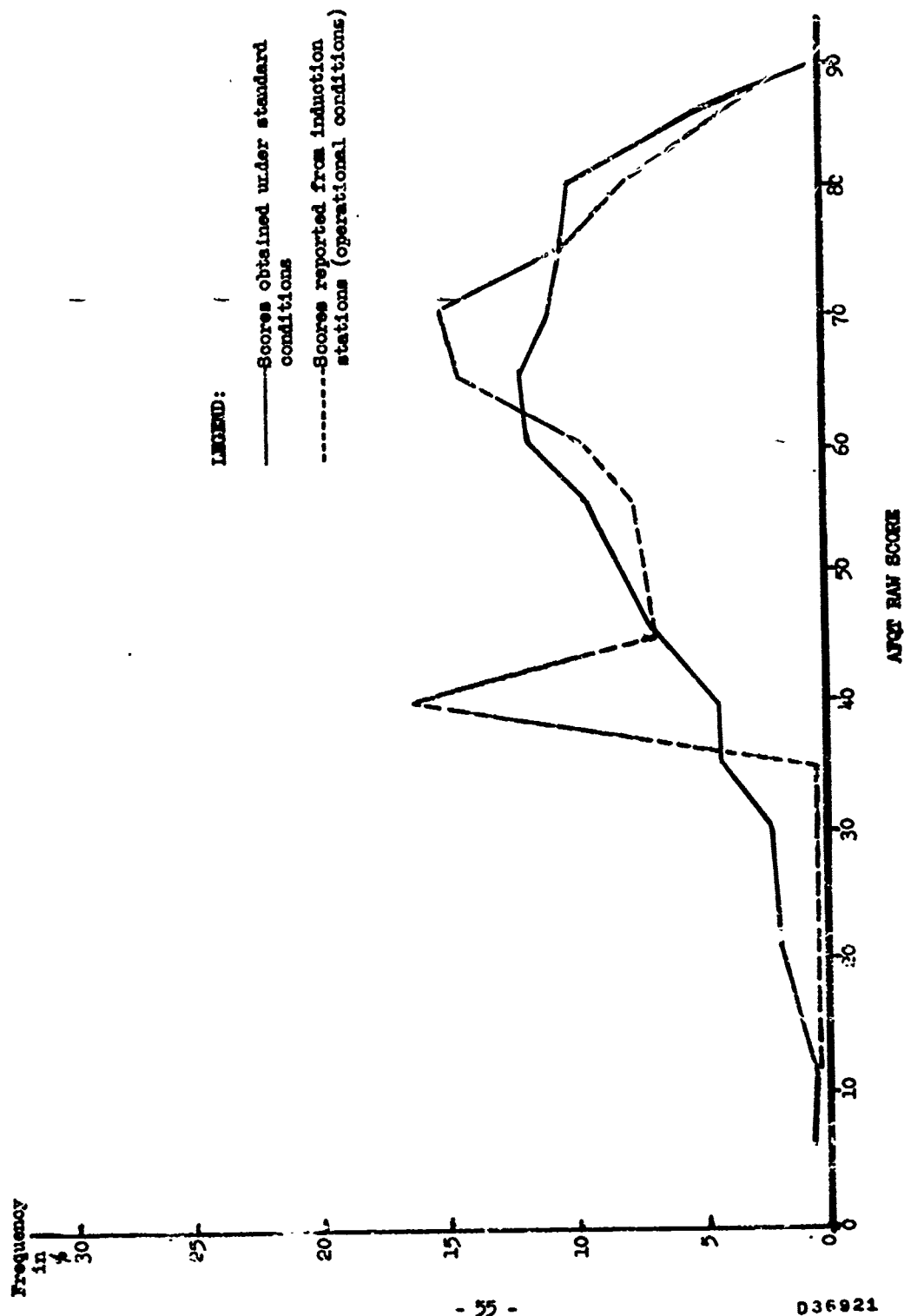Figure J-1. AFQT scores of same enlistees from operational and standard administrations (N = 288).

LEGEND:

———— Scores obtained under standard conditions

——————— Scores reported from induction stations (operational conditions)

AFQT RAW SCORE

Figure J-2. AFQT scores of same inductees from operational and standard administrations (N = 712).

Frequency in %

D36921